UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Ciências
ULisboa

Evolutionary history of sympatric Rainbow skinks from the Australian Monsoonal Tropics

Doutoramento em Biologia

Especialidade em Biologia Evolutiva

Ana Catarina Afonso Silva

Tese orientada por:

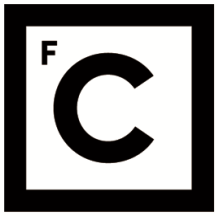
Professor Doutor Craig Moritz, Professora Doutora Maria Manuela Gomes Coelho Noronha Trancoso e Professor Doutor Carlos Fernandes

Documento especialmente elaborado para a obtenção do grau de doutor

2018

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS



Ciências
ULisboa

Evolutionary history of sympatric Rainbow skinks from the Australian Monsoonal Tropics

Doutoramento em Biologia

Especialidade em Biologia Evolutiva

Ana Catarina Afonso Silva

Tese orientada por:

Professor Doutor Craig Moritz, Professora Doutora Maria Manuela Gomes Coelho Noronha Trancoso e Professor Doutor Carlos Fernandes

Júri:

Presidente:

- Doutor Henrique Manuel Roque Nogueira Cabral, Professor Catedrático da Faculdade de Ciências da Universidade de Lisboa

Vogais:

- Doutor Craig Moritz, Professor da Australian National University
- Doutor Nuno Miguel dos Santos Ferrand de Almeida, Professor Catedrático da Faculdade de Ciências da Universidade do Porto
- Doutor Paulo Célio Pereira Martins Alves, Professor Associado da Faculdade de Ciências da Universidade do Porto
- Doutora Joana Isabel do Espírito Santo Robalo, Professora Auxiliar do Instituto Universitário de Ciências Psicológicas, Sociais e da Vida
- Doutora Margarida Maria Demony de Carneiro Pacheco de Matos, Professora Associada com Agregação da Faculdade de Ciências da Universidade de Lisboa
- Doutor Vítor Martins Conde e Sousa, Professor Auxiliar Convidado da Faculdade de Ciências da Universidade de Lisboa

Documento especialmente elaborado para a obtenção do grau de doutor

Fundação para a Ciência e a Tecnologia (FCT) – Bolsa de Doutoramento (SFRH/BD/88740/2012)

2018

The work presented in this dissertation was mainly developed at Division of Ecology and Evolution, Research School of Biology at The Australian National University (Canberra, Australia) and at the Departamento de Biologia Animal da Faculdade de Ciências da Universidade de Lisboa, with support from Centre for Ecology, Evolution and Environmental Changes (CE3C).

This study was supported by Fundação para a Ciência e a Tecnologia (FCT), through a PhD grant (SFRH/BD/88740/2012), attributed to Ana Catarina Afonso Silva.

This dissertation should be cited as:

Afonso Silva, A. C. (2018) Evolutionary history of sympatric Rainbow skinks from the Australian Monsoonal Tropics. PhD Thesis, University of Lisbon, Portugal.

Preliminary Note

The present dissertation includes original scientific papers already published (3), or in submission process (1) to international indexed journals. According to Article 31 nr.1 of the Post-Graduate Studies Regulation (Diário da República, 2.ª série, n.º 57, de 23 de Março de 2015,) the PhD candidate declares full participation in the scientific planning, data collection, analyses, discussion and writing of all manuscripts presented here as first author. In the manuscript where the PhD candidate participated, the contribution was on the design of research, data collection, discussion and participation on the writing of the manuscript.

The dissertation, being composed of a series of international publications, is written in English. A reference list is given at the end of each chapter rather than at the end of the thesis. Due to this format, there may be instances of duplication across chapters. Each chapter contains its own Supplementary material where required. The different format of some chapters reflects the specific requirements of the scientific journals to which the presented manuscripts were submitted.

Nota Prévia

Na elaboração desta dissertação foram usados integralmente artigos científicos já publicados (3) ou em fase de submissão (1) em revistas científicas indexadas. De acordo com o previsto no nº 1 do artigo 31 do Regulamento de Estudos Pós-Graduados da Universidade de Lisboa, publicado no Diário da República, 2.ª série, n.º 57, de 23 de Março de 2015, a candidata esclarece que participou na conceção, obtenção dos dados, análise, discussão dos resultados e redação dos respetivos manuscritos dos trabalhos em que foi primeira autora. No trabalho em que a candidata colaborou, a contribuição foi na conceção, obtenção de dados, discussão de resultados e participação na escrita.

A dissertação, por ser uma compilação de publicações internacionais, está redigida em Inglês. Uma lista de referências é dada no final de cada capítulo em vez de no final da tese e devido a este formato poderá haver casos de duplicação entre capítulos. Cada capítulo contém também a informação de suporte a ele associada. O formato diferente de alguns capítulos reflete os requerimentos específicos dos jornais científicos aos quais os manuscritos apresentados foram submetidos.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

To get to the end of this thesis it took literally a village and many people need to be acknowledged for their support during the years it took to get here.

First, I would like to thank my main supervisor and mentor, Prof. Craig Moritz, for all his support since the beginning. The opportunity to work with him was by far one of the main outcomes from my PhD, his scientific enthusiasm was a big driver to keep me going and I hope to be able to apply throughout my career the advice I received from his mentorship. I would like to also thank to my other supervisors, Prof. Manuela Coelho and prof. Carlos Fernandes, for supporting my research career since the beginning during the degree and master time, all your support took me to Australia and back from there, and this was very important right until the end.

Thanks to the various museums in Australia (MAGNT, WAM, SAM, MV, QM, AM, ANWC) that provided the samples and tissues for this work, specifically to Paul Doughty and Ryan Ellis for their support with specimens and taxonomy. To Traditional Owners of Country, National Parks Australia and private properties that were also important during the various field seasons, in particular to Cecilia Myers from the Dunkeld Pastoral Company for all the access and resources during field work and experiments in Theda. Additionally, thanks to Henry Cook, Brendan Schembri, Russell Barrett, Stewart Macdonald and Stephen Zozaya for all the photos and natural history notes on rainbow skinks.

I would like to thank to everyone from the Moritz Lab, for their academic and personal support during my PhD, specially to my close collaborators Sally Potter, Jason Bragg and Huw Ogilvie for their time, patience and motivation to push the research I present here forward. Also thanks to Richard Harrison, Dan Rosauer, Maxine Piggott, Renae Pratt, Renee Catullo, Paul Oliver, Nick Matzke, Eric Rittmeyer, Maria Strangas, Luisa Teasdale, Catherine Noble, Lauren Ashmann, Rosa Agudo, Felipe Martins, Maggie Leigh, Leonardo Tedeschi, for support, discussions during lab meetings and in other social events. Special thanks to Mitchell Scott, Bevan Yiu, Gaye Bourke, Mitzy Pepper, Martha Muñoz, Mark Allen, that with Craig and Sally provided great moments and experiences during the four field seasons I participated in the amazing Australian monsoon tropics. Also thanks to previous Moritz Lab students, specially Charlotte Jennings, Matt Fujita and Sonal Singhal for various discussions and advice. Thanks to Renata Gonçalves and Natali Santos for support with measurements of specimens. Also thanks to Emma Sherratt, Damien Esquerré, Ian Brennan, Liam Bailey and Marta Vidal-Garcia that provided for assistance with several statistical analyses.

A very special thanks to the several PhD sisters and brothers I had the luck to have and that have been there for me every time I needed, either to discuss science or just to have some good fun: Christiana McDonald-Spicer, Jessica Antunes, Becky Laver, Huw Ogilvie, and specifically Mozes Blom and Joshua Peñalba. Without these two guys a lot more could have gone wrong, they were, still are and hopefully will continue to be, my anchor that are always there when I need, to make me laugh or just to listen to my complaining, in scientific discussions or just for gossip, my brotherhood with these two guys is another great achievement during my time in Australia.

Furthermore, I had the opportunity to get to know a lot of wonderful people that made my PhD time much more enjoyable and whom I am very thankful for their friendship and good times in Canberra. Thanks to Nina, Robyn and David, that together with Becky

and Josh had to endure the stressful time in the writing PhD office. Thanks to Daniela, Jessie, Sonya and Bo for our PhD task force that kept me going in the last year. Thanks to everyone part of the bootcamp group that kept me active and sane, specially Jessie, Pip and Zoe. Also thanks to Catherine & Dave, Anna, Iliana, Coni, Carlos, Stu, Jono & Karo, Hee Jin, Frances, Jared, Hannah, Danswell, Hoops, Thomas Merkling, Thomas Wallenius, Megan Head, Megan & Florian, Alyssa, Tom, Dan Wait, Rita, Alex, Kenneth, Kristal, Foteini, and Lalo and Regina, for all the really enjoyable time in Canberra. Everyone in ANU really received me as if I was an official student in ANU, which was particularly special that when everyone came to celebrate with me the time I spent there, I was surprised by everyone contributing to help me with my tuitions. I am very grateful for the community feeling I was able to experience in ANU and for all the friendships that still are with me, especially with Jessie, Liam, Josh, Moos and Regina.

Também preciso de agradecer a todos que em Portugal me apoiaram quando decidi ir o outro lado do mundo para fazer ciência e andar atrás de uns lagartos. Obrigada ao André e à Raquel, a todos com quem partilhei e discuti ideias do grupo do Carlos, e da prof. Manuela, em especial o apoio do Vítor e Mónica nos últimos meses.

Obrigada às minhas meninas que sempre estiveram lá comigo desde os tempos do básico e mesmo à distância, à Daniela, Dili, Joana, em especial à Inês por ajudar-me a rever o português na tese e à Yola por ter disponibilizado a casa que deu a força final que precisava para acabar este desafio. Também um obrigado muito especial à minha família, que mesmo sem perceber porque quis seguir este percurso tem estado sempre lá a apoiar-me, em especial a minha avó e a minha mãe.

ABSTRACT

How speciation, range shifting and reticulation through climatic oscillations combine to shape current patterns of local and regional diversity remains a key question in evolutionary biology. This can be investigated by using multi-scale analyses of closely related species assemblages in diverse communities, such as the lizards of Australia's tropical savannas. An example of these is the understudied group of rainbow skinks (*Carlia*) that are broadly co-distributed across the Australian Monsoonal Tropics (AMT) region, and that stands out from most well-studied *Carlia* species occurring along the Australian east coast.

The aims of this thesis were to: i) identify lineage diversity between two sympatric and closely related skinks and to use an integrative taxonomic approach to statistically test major lineages as species; ii) compare past responses of species with different climatic niche breadth that experienced the same climatic fluctuations; iii) infer the phylogenetic relationships of the rainbow skinks in a collaborative study; and iv) explore the occurrence of introgression in a group of six species that broadly co-occur in the AMT, using a dataset by target exon capture (>1000 loci).

The investigation of lineage diversity discovered cryptic lineages mostly in the Kimberley region (north west of Australia). In an integrative taxonomic approach, these lineages were then statistically validated as new species using multispecies coalescent methods and morphological analyses, and subsequently described as *C. insularis* sp. nov and *C. isostriacantha* sp. nov. with genetic and morphological characters. The exploration of how concordant were the species responses with late Pleistocene climatic changes, identified contrasting responses by species with different climatic niches, suggesting that the narrow climatic specialist species was more sensitive to these changes.

In addition, using multiple phylogenomic approaches, it contributed to a much improved and well-supported phylogeny for the rainbow skinks in comparison with a previous, poorly resolved tree. The new tree showed that the six *Carlia* species that co-occur in the AMT are actually closely related.

Lastly, the analysis for the presence of introgression between the AMT sympatric species did not detect evidence of recent admixture, but identified patterns of ancestral introgression before the divergence of sister species, and some instances of introgression in the more climatic unstable Kimberley region. It also showed that by accounting for reticulated evolution with phylogenetic networks methods, a distinct topology from the overall well supported species tree can be observed, in this case with the clade inferred as introgressed appearing as more ancestral in the phylogeny.

The results of this thesis have implications for the understanding of the processes driving cryptic species diversity and responses to past climatic change across this rich and understudied Australian Monsoonal Tropics biome.

Keywords: phylogeography, integrative taxonomy, introgression, *Carlia*, exon capture sequencing

RESUMO

A Biologia Evolutiva tem vindo a contribuir significativamente para o conhecimento dos efeitos das oscilações climáticas passadas e presentes na formação dos atuais padrões de diversidade regional e local, em particular nos processos de especiação e reticulação, bem como na alteração da distribuição de espécies. Esta questão pode ser investigada através de análises a várias escalas espaciais e temporais, tais como num conjunto de espécies relacionadas e provenientes de comunidades diversas, como é o caso dos lagartos que ocorrem nas savanas tropicais do norte da Austrália. O presente estudo aborda um grupo de espécies de lagartos pouco estudados, os escincídeos do género *Carlia* que se encontram co-distribuídos na região das monções tropicais da Austrália ("Australian monsoonal tropics", AMT). As espécies desta região distinguem-se de outros escincídeos *Carlia* que ocorrem ao longo da costa leste do continente e que têm sido alvo de vários estudos.

Os objetivos desta tese foram: i) identificar linhagens entre duas espécies simpátricas e próximas de escincídeos e usar uma abordagem de taxonomia integrativa para testar estatisticamente as linhagens principais como espécies separadas; ii) comparar respostas a flutuações climáticas no passado de espécies com diferentes nichos climáticos; iii) num estudo colaborativo, inferir relações filogenéticas dos escincídeos arco-íris; e iv) explorar a ocorrência de introgressão num grupo de seis espécies que coexistem na região AMT, usando *loci* obtidos por sequenciação de exões obtidos pelo método de captura direcionada.

É expectável que espécies endémicas de regiões tropicais sejam vulneráveis a alterações climáticas no futuro, em parte devido aos seus nichos climáticos mais restritos. Do mesmo modo, é também provável que estas espécies exibam padrões filogeográficos acentuados resultantes da sua resposta a alterações climáticas no passado. Para testar a hipótese de que espécies especialistas tropicais são mais sensíveis a alterações climáticas do que espécies generalistas em termos climáticos, foram gerados e analisados dados baseados no ADN mitocondrial e em mais de 2000 *loci* resultantes da sequenciação de exões obtidos pelo método de captura direcionada, com o objetivo de comparar padrões de persistência histórica e de flutuações populacionais em duas espécies próximas de *Carlia*: *C. johnstonei* (especialista tropical) e *C. triacantha* (generalista climática). De acordo com a generalidade dos trabalhos existentes, previa-se que a espécie especialista tropical apresentasse mais estrutura e flutuações demográficas mais pronunciadas do que a espécie generalista climática, que por sua vez deverá ter tido populações mais estáveis durante períodos de clima mais severo no Quaternário Tardio. Verificou-se que para *C. johnstonei*, as

populações de ilhas da região de Kimberley, no Noroeste da Austrália, são fortemente divergentes das populações continentais e de outras ilhas próximas. Em *C. triacantha*, um dos clados principais ocorre desde os desertos no centro da Austrália até à região mais mésica (habitat com humidade moderada) do Top End (norte central da Austrália) e um outro clado ocorre principalmente em Kimberley, mas também em outras localidades do centro da Austrália. Na região de Kimberley, onde a distribuição destas linhagens se sobrepõe, os dados genéticos sugerem a existência de estrutura filogeográfica e expansão demográfica mais forte para a espécie especialista tropical em comparação com a espécie mais generalista. Estes resultados são consistentes com a hipótese de que especialistas tropicais terão sido mais sensíveis a alterações climáticas do passado.

Embora os métodos genéticos para delimitação de espécies tenham vindo a ser visivelmente melhorados durante a última década, esta área continua em desenvolvimento e expansão. Idealmente, abordagens com dados genéticos e baseadas em modelação devem ser consideradas juntamente com outras evidências, tais como morfológicas e geográficas, numa abordagem de taxonomia integrativa. Fortes divergências filogeográficas têm sido descritas para várias espécies de lagartos do género *Carlia*, mas os estudos anteriores focaram-se apenas em algumas linhagens da costa leste da Austrália. No presente estudo, para além de terem sido descobertas diferentes linhagens filogeográficas em diferentes espécies do noroeste da Austrália, foi possível apresentar uma validação estatística destas linhagens. Foi também introduzida uma nova abordagem, baseada num método recentemente disponível no software "StarBEAST2", que assenta no modelo "Coalescente multiespécies" (CMS) e usa múltiplos loci e fatores Bayes (BFD) para explorar o suporte de hipóteses alternativas de delimitações de espécies. Este método foi aplicado juntamente com outros dois métodos (BFD* SNAPP e BPP) baseados no mesmo modelo, usando os mesmos dados de sequenciação de exões. Foram também utilizadas 11 medidas morfológicas numa abordagem integrativa para avaliar evidência estatística da diferenciação morfológica das duas linhagens candidatas a espécies. Os resultados dos três métodos baseados na coalescência e na morfologia suportaram a diferenciação de nível específico das quatro linhagens, duas em cada uma das espécies originais. A abordagem introduzida neste estudo (BFD StarBEAST2) produziu resultados bastante congruentes com os outros métodos tradicionalmente utilizados em delimitação de espécies, suportando o uso desta nova abordagem para validação estatística de espécies. Tal como observado com a divergência genética, mais pronunciada em *C. johnstonei*, a divergência morfológica foi superior entre as linhagens desta espé-

cie quando comparada com as linhagens de *C. triacantha*. Esta congruência proporcionou o reconhecimento de duas espécies dentro de *C. johnstonei* com forte suporte, e todas as evidências combinadas também sustentaram a existência de duas espécies dentro de *C. triacantha* ao longo da sua extensa distribuição geográfica. Assim, este estudo compreende a descrição de duas espécies novas, *Carlia insularis* sp. nov. e *Carlia isostriacantha* sp. nov., ambas do Noroeste da Austrália, o que contribuiu para sublinhar a riqueza da fauna desta região tropical da Austrália.

A aplicação de sequenciação de nova geração utilizando captura direcionada tem recentemente permitido análises filogenómicas de clados em que a especiação terá ocorrido por radiação rápida. Este tipo de análises pode ser dificultado na presença de retenção de muito polimorfismo ancestral, o que exige o uso de métodos que têm explicitamente em conta este fenómeno, como são exemplo os métodos baseados no modelo de CMS. Tais dificuldades metodológicas foram exploradas num estudo colaborativo, baseado também na sequenciação direcionada a exões capturados, que teve por objetivo inferir com robustez as relações filogenéticas das espécies Australianas do género *Carlia* e de outros géneros relacionados que no seu conjunto representam os escincídeos arco-íris. Análises filogenéticas anteriores realizadas neste grupo de espécies revelaram pequenas distâncias genéticas entre clados e discordância significativa entre *loci*, o que limitou a resolução da inferência da filogenia de espécies num grupo que terá diversificado rapidamente. Neste estudo foram obtidas filogenias através de métodos que têm a capacidade de utilizar todos os *loci* disponíveis, quer através da concatenação dos *loci* ou de abordagens sumárias baseadas no CMS que utilizam filogenias de genes para obter uma filogenia de espécies final. Ambas as abordagens inferiram topologias relativamente consistentes e bem resolvidas, incluindo para clados mais ancestrais que em estudos anteriores apresentaram afinidades ambíguas. Adicionalmente, testes realizados com diferentes estratégias para a amostragem de haplótipos em cada espécie com estrutura filogeográfica para as análises de CMS, mostraram que a combinação de indivíduos de diferentes linhagens filogeográficas consegue produzir filogenias de espécies consistentes. Este estudo também mostrou que, contrariamente ao sugerido pela filogenia que emergiu dos estudos anteriores, as espécies que ocorrem na região AMT (*C. amax, C. rufilatus, C. munda, C. gracilis, C. johnstonei e C. triacantha*) são espécies próximas dentro do mesmo clado filogenético.

O aumento da disponibilidade de dados de sequenciação na era da genómica tem mostrado que a ocorrência de hibridação introgressiva entre espécies em divergência é mais comum do que o esperado. Novas técnicas com a capacidade de dis-

tinguir entre a retenção de polimorfismo ancestral e eventos de introgressão no passado têm também sido desenvolvidas, permitindo o estudo do papel da introgressão, na história evolutiva das espécies, bem como na persistência de espécies em ambientes passados instáveis. Neste estudo foram investigados padrões de introgressão em seis espécies simpátricas de escincídeos *Carlia* que ocorrem na região AMT, e que o estudo filogenético acima referido mostrou serem espécies próximas. A partir de dados de sequenciação direcionada a exões capturados, foi possível testar a eventual ocorrência de mistura genética recente, bem como evidências de introgressão antiga, entre espécies e entre linhagens filogeográficas simpátricas em regiões com diferentes histórias climáticas. Apesar da ampla sobreposição geográfica entre espécies e linhagens, apenas foi pontualmente detetada a possibilidade de mistura recente, o que sugere forte isolamento reprodutivo entre espécies. No entanto, foi obtido suporte para a ocorrência de vários eventos de introgressão no passado. Apesar de não terem sido detetados casos de introgressão entre linhagens filogeográficas de espécies diferentes que pudessem sugerir hibridação mais recente, as ocorrências de introgressão regional detetadas sugerem um potencial papel da hibridação em regiões de instabilidade climática elevada. Este estudo mostrou como a introgressão pode potencialmente ter consequências significativas na história evolutiva de espécies, como por exemplo durante o processo de especiação ou na capacidade de resposta das espécies na presença de instabilidade ambiental no passado.

No geral, os resultados deste estudo contribuem para uma melhor compreensão dos processos que originam diversidade críptica, bem como das respostas das espécies a alterações climáticas no passado, na região bastante rica e pouco estudada do bioma Australiano de Monções Tropicais.

Palavras-chave: filogeografia, taxonomia integrativa, introgressão, *Carlia*, sequenciação direcionada a exões capturados

LIST OF ABBREVIATIONS

θ_η	Theta per site from total number of mutations
1S	Experimental treatment – alleles from one sample
2S1L	Experimental treatment – alleles from 2 samples from 1 lineage
2S2L	Experimental treatment – alleles from 2 samples from 2 lineages
ABC	Approximate Bayesian Computation
ABTC	Australian Biological Tissue Collection at the South Australian Museum
AGL	Axilla-groin length
AHE	Anchored hybrid enrichment
AIC	Akaike Information Criterion
AMT	Australian Monsoonal Tropics
ANU	Australian National University
AWT	Australian Wet Tropics
BF	Bayes factor
BFD	Bayes Factor Delimitation
BIC	Bayesian Information Criterion
bp	Base pairs
CCM	Craig Moritz lab collection
CIL	Complete and informative loci
CL	Complete loci
Da	Net sequence divergence
DNA	Deoxyribonucleic acid
Dxy	Absolute sequence divergences
EAL	Ear aperture length
EC	Exon capture
EED	Eye to ear distance
FLL	Forelimb length
F_{st}	Fixation Index
GQ	Genotype quality score
HD	Head depth
H_d	Haplotype diversity
HL	Head length
HLL	Hindlimb length
HW	Head width
ILS	Incomplete lineage sorting
IM	Isolation-with-migration
k	Average number of nucleotide differences
km	Kilometer
LDA	Linear Discriminant Analysis
LGM	Last Glacial Maximum
MAGNT	Museum and Art Gallery of the Northern Territory
MANOVA	Multivariate Analysis of Variance
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood

mm	Millimetres
MSC	Multispecies Coalescent
mtDNA	Mitochondrial DNA
MV	Museum Victoria
MYA	Million years ago
n	Sample size
ND4	NADH dehydrogenase subunit 4
nDNA	Nuclear DNA
NS	Nasals separation
NT	Northern Territory
PC	Principal Component
PCA	Principal Components Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PDL	Palpebral disc length
Psi	Directionality index
RF	Random Forest Analysis
RI	Reproductive Isolation
s. l.	*sensu lato*
s. s.	*sensu stricto*
SAM	South Australian Museum
SFS	Site Frequency Spectrum
SI	Strict isolation
SNP	Single Nucleotide Polymorphisms
Sp	Species
sp. nov	*species nova*
SVL	Snout-vent length
TC	Tree Certainty
UCE	Ultraconserved elements
WAM	Western Australian Museum
η	Total number of mutations
θ_S	Segregating sites
π	Nucleotide diversity

Chapter I

General Introduction

1.1 Species evolutionary histories and responses to climate change

Studying species' evolutionary histories requires an understanding at different scales in the speciation continuum from population genetics, phylogeography, to phylogenetics and spanning studies working within or above the species level (Cutter 2013). Specifically, phylogeography is an important field as it bridges the understanding of intra- and interspecific patterns (Avise et al. 1987, 2016), and is a route to detecting cryptic species, integrating geography and environmental change into the study of lineage diversification, and investigating historical processes of community assembly (Hickerson et al. 2010).

1.1.1 Phylogeography

The field of phylogeography has seen much progress since it was first coined by Avise et al. (1987), with a recent issue reviewing much of the state of the field and summarized in Avise et al. (2016). From the beginning of using cytoplasmic loci, that for the longest time provided the bulk of species specific empirical studies, to the use of multiple nuclear markers (Brito & Edwards 2009) and hypothesis-driven studies (Knowles 2004), phylogeography is becoming much more integrative with other approaches (such as landscape genetics and paleoecology, Rissler 2016; Prates et al. 2016). Current phylogeographic studies are much more comparative using multiple species to understand general biogeographic patterns (Bowen et al. 2016), incorporate more trait data (Zamudio et al. 2016) with consequences for integrative taxonomy, applying genomic datasets (McCormack et al. 2013) and coalescent-based methods to infer the most likely scenario by estimating the probability of multiple demographic models (Thomé & Carstens 2016).

Phylogeography has been very successful in inferring congruent patterns within different biomes starting with well-studied biotas from western North America (Brunsfeld et al. 2001; Cook et al. 2001; Shafer et al. 2010), south-eastern North America (Avise 2000; Soltis et al. 2006) and Europe (Taberlet et al. 1998; Hewitt 1999, 2001; Weiss & Ferrand 2007; Feliner 2011; Nieto Feliner 2014). Although southern hemisphere and other regions in developing countries have taken longer to provide similar studies (Beheregaray 2008), several reviews have been showing recent progress such like in African rain forest and Savannas (Lorenzen et al. 2012; Hardy et al. 2013), South America savannas and open vegetation biomes (Werneck 2011; Turchetto-Zolet et al. 2013), Tibetan biota (Favre et al. 2015), temperate and arid China (Qiu et al. 2011; Meng et al. 2015), Indo-Australian archipelago (Lohman et al. 2011), more specifically the Sunda shelf (Lohman et al. 2011) and several biomes within Australia (Schneider & Moritz 1999; Byrne et al. 2008, 2011; Byrne 2008; Bowman et al. 2010).

Much of the focus of phylogeography has been on understanding the effects of late Quaternary events on species current distributions (Hewitt 1996, 2011). The climatic oscillations of the Quaternary influenced the geographical ranges of species to expand and contract in a cyclical manner (Hewitt 1996). However, how individual species respond to climate

change depends largely on their intrinsic traits (such as specific adaptations and environmental tolerances), whether they will either adapt to the new environment or geographically shift their range to track their climatic niche; and if not capable of either, decline to the point where they might go extinct (Davis et al. 2005).

Comparative phylogeographic studies can identify regions in which multiple taxa persisted through past climate change in biogeographic refugia (Davis et al. 2005). The duration of isolation in these refugia and the mode of recovery after instability can reflect how species' responded to climate change (Davis et al. 2005), whereas a low consequence outcome would be populations isolated for a short period in different refugia, then coming back into contact, and due to low divergence, migration between populations would homogenise across the species the independent diversity that arose in the refugia. This in contrast with a more detectable outcome observed in current parapatric species, which is consistent with the reduction of populations' range to survive in refugial regions for a long period of time, in which expansion after the instability would form contact zones between well differentiated species, possibly with subsequent hybridization and introgression (Hewitt 2011). Understanding how species have responded to past climate changes, and whether they endured periods of adverse climates, can be used in models forecasting how current and future climate change may affect species capacity to adapt, move or endure the predicted instability (Espíndola et al. 2012; Moritz & Agudo 2013). However, multiple studies suggest disparate species' responses to climate change for sympatric species, with some capable of widespread persistence, while others persist in just a few major refugia and this despite a common history of climate change in the same region (Byrne 2008; Moussalli et al. 2009; Bell et al. 2010; Prates et al. 2016).

1.1.2 Predictions of generalist vs specialist responses to climate change

The tendency of species to retain ancestral ecological characteristics over time (Wiens & Graham 2005) is referred as niche conservatism. Although this seems to be a common pattern (Wiens 2004), should closely related species have different niches, then there is opportunity to better understand differential responses to climate change.

The niche breadth hypothesis suggests species encountering a greater set of climate conditions across their range (generalist species) are expected to have broader tolerances to climate change and larger geographic ranges than climatically restricted species (specialist species) (Brown 1984, 1995). In addition, it is expected that species with restricted geographic ranges are more likely to be sensitive to climate change than widespread species (Johnson 1998). Species with restricted ranges and narrower niche breadths are often considered to have lower dispersal capacity, be weak competitors and have small local populations (Gaston 1994), which would compromise the capability of species with restricted ranges to extend their distribution to new habitats (Thuiller et al. 2005).

Related to the niche variation hypothesis, Janzen (1967) predicted that tropical ectotherms will be more thermal specialists in comparison with higher-latitude relatives, because

they would have evolved in relatively constant thermal environments with reduced seasonality. And following this hypothesis, even within a temperate region, Chejanovski & Wiens (2014), found that areas with higher species richness had species with narrower climatic niches. Ghalambor (2006) summarised how these niche breadth and dispersal predictions relate to increase opportunities for allopatric speciation, and (Slatyer et al. 2013) also found support for a general positive relationship between niche breadth and geographic range.

Walters et al. (2012) suggests the prediction that the increase in thermal specialization towards the tropics (Tewksbury et al. 2008) will lead to stronger selection to reduce genetic variance (Huey & Kingsolver 1993; Angilletta Jr 2009). This suggests a trade-off between the maximum and breadth of performance, where thermal specialists are favoured over thermal generalists in a rapidly changing environment, while if genetic variation increases with increasing breadth, then populations of thermal generalists will tolerate the greatest rates of environmental change (Huey & Kingsolver 1993).

1.2 Phylogenetics and species delimitation in the age of genomics

1.2.1 NGS markers and advantages of sequence capture

In the age of genomics and massively parallel sequencing, several approaches have been developed to more readily assess genome diversity and divergence either at inter- or intra-specific levels. While the use of whole-genome datasets to solve evolutionary relationships have mostly been applied to well-studied groups (e.g. Drosophila, Clark et al. 2007; birds, Jarvis et al. 2014, and mammals, Sims et al. 2009) most phylogenetic and phylogeographic studies have used either SNP-based methods like RADseq (review of methods in Puritz et al. 2014) or sequenced-based target capture approaches (reviewed in Lemmon & Lemmon 2013; Jones & Good 2016). Due to short fragment size and patchy recovery of homologous loci at deeper phylogenetic scales (Rubin et al. 2012; Leaché & Oaks 2017), SNP-based methods are mostly used to analyse genetic diversity between populations and relatively closely related species (McCormack et al. 2013). By contrast target capture methods can be applied to generate high quality and relatively complete data within species (Bi et al. 2012; Brandley et al. 2015) and across species at larger phylogenetic scales (Jones & Good 2016).

Recent efforts have been made to develop sequence capture sets that can be used across a wide level of phylogenetic scales, as broad as UCEs (ultraconserved elements) for Amniota (Faircloth et al. 2012), anchored hybrid enrichment sets (AHE) for vertebrates (Lemmon et al. 2012), exon sequence capture for the invertebrate class Ophiuroidea (Hugall et al. 2016) and marsupials (Bragg et al. 2017), all squamate reptiles (Singhal et al. 2017) or more specifically for a given clade, e.g. Skinks (Bragg et al. 2016).

The main impediment to the widespread use of targeted capture with non-model species are lower number of retrieved loci, the starting costs associated with designing a capture probe set and the need of a priori knowledge of target sequences (Elshire et al. 2011). How-

ever, these methods yield data assembled into individual longer fragments, resulting in haplotypes or consensus sequences useful for phylogenetic analysis, while still capable of extracting hundreds to thousands of biallelic SNPs suitable for population genetics. Target capture sets can be developed from transcriptomes for the clade in question, with the advantage to be applied to degraded DNA, like museum specimens (Bi et al. 2012), and are effective across moderate levels of phylogenetic divergence (Bragg et al. 2016; Hugall et al. 2016), while still useful to detect intraspecific variation (Potter et al. 2016). While UCEs, AHE and exon sequence capture can all be used to retrieve targeted sequence data, customize exon capture probes can be designed with suitable length to target higher diversity exons than the conserved loci from the other two approaches (Brandley et al. 2015; Jones & Good 2016).

1.2.2 New methods for phylogenetic inference

The increase in available genetic data has brought the need for better and faster methods to assess species evolutionary histories. For phylogenetic analysis of hundreds or thousands of loci, most researchers have concatenated loci into a super matrix and typically obtain well supported phylogenies using maximum likelihood methods like RAxML (Stamatakis 2014) or IQ-TREE (Nguyen et al. 2015). However, concatenation overestimates confidence of relationships and with rapid divergences (relative to population size) will incorrectly reflect evolutionary relationships (Kubatko & Degnan 2007). This is because gene tree discordance due to ILS is not accounted for with concatenation, which assumes that all gene trees have the same underlying topology. Failure to account for coalescence processes will also upwardly bias estimates of branch length towards the tips (Ogilvie et al. 2016; Mendes & Hahn 2016) which could have negative implications for subsequent inference of macroevolutionary patterns.

Recent coalescent-based methods have been developed to explicitly take into account ILS as a source of gene tree discordance (Degnan & Rosenberg 2009) to estimate species trees (reviewed in Liu et al. 2015). These methods are consistent under the multispecies coalescent model, i.e. they will converge in probability to the true species tree with the increase in number of loci (Warnow 2015); and can be generally classified into summary statistic and full Bayesian methods.

Summary statistic methods estimate the species tree from a set of gene trees, which are often being estimated using maximum likelihood methods. With this approach, it is easier to parallelize gene tree estimation and species tree methods such as ASTRAL (Mirarab et al. 2014), ASTRID (Vachaspati & Warnow 2015) or MP-EST (Liu et al. 2010) are generally very fast for many loci and taxa (Liu et al. 2015a). Despite their appeal, these methods do have limitations, including that they do not directly estimate time trees and are highly sensitive to violations of the underlying multi-species coalescent model, including introgression (Solís-Lemus et al. 2016).

Bayesian coalescent methods make more efficient use of data by jointly estimating species tree and gene trees within species trees directly from multilocus sequence data. These methods use the posterior probability distribution approximated by sampling trees generated from a Markov Chain Monte Carlo (MCMC) algorithm (Liu et al. 2015a). Consequently, methods like BEST (Edwards et al. 2007; Liu & Pearl 2007; Liu 2008), *BEAST (Heled & Drummond 2010) and BPP (Zhang et al. 2014) have high computational costs (Liu et al. 2015a). Recently some improvements have been made under the Bayesian multispecies coalescent methods, like the use of biallelic SNPs with SNAPP (Bryant et al. 2012), and algorithmic improvements to increase the efficiency of computation in BPP (Rannala & Yang 2017) and StarBeast2 (Ogilvie et al. 2017).

1.2.3 Species delimitation

Species delimitation can be a challenging task considering the difficulty to decide which evidences are sufficient to define and diagnose species, considering the debate around of more than 30 species concepts (de Queiroz, 1998). Here, this thesis follows the unified species concept (De Queiroz, 2007) that defines a species by treating existence as a separately evolving metapopulation lineage as the only necessary property of a species, with different lines of evidence from the more applicable methodologies in assessing lineage separation. Also, delimiting species should not be considered as a final static outcome, since species are hypotheses that are subject to change with new data and methods (Fujita et al. 2012).

The proliferation of genetic markers and phylogenetic methods ultimately benefited the field of species delimitation, the main purpose of which is to identify species-level biological diversity (Carstens et al. 2013). Genetic species delimitation methods consist of discovery approaches, which assign samples to groups without any a priori information, and validation approaches that are based on hypothesis testing and require a prior assignment of samples of lineages (O'Meara 2010; Carstens et al. 2013).

Most species delimitation approaches tend to only perform discovery analyses while investigating genetic substructure in populations (Rannala 2015), using SNP-based clustering methods like Structure (Pritchard et al. 2000) or Structurama (Huelsenbeck et al. 2011), or gene tree-based methods like GMYC (Pons et al. 2006) or Brownie (O'Meara 2010). However, with many loci very subtle differences between populations can be detected, even if low ongoing gene flow still occurs, requiring further evaluation of identified lineages with additional analyses (Rittmeyer & Austin 2012; Carstens et al. 2013). Hence these clustering approaches are not, by themselves, useful for species delimitation (Rannala 2015).

The most commonly used validation approaches are based on the multispecies coalescent model, since it provides clear and objective testing of alternative hypothesis for independent lineages (review in Fujita et al. 2012). A Bayes Factor (BF) approach was implemented with the SNP-based SNAPP (Leaché et al. 2014) to calculate marginal likelihoods for use in comparing BFs of particular delimitation hypothesis. Although the advantage of

SNAPP is that it can integrate over many gene trees, the BF implementation relies on a computationally intensive MCMC path sampling framework (Rannala 2015). In contrast, BPP uses a reversible-jump MCMC algorithm to estimate the posterior probability of alternative delimitation hypotheses, and this can be performed jointly with inferring the species tree (Yang & Rannala 2014).

Carstens et al. (2013) recommended that multiple delimitation methods should be applied and delimitation should be accepted when it is the most common across most of the methods. However, from a statistical perspective, Rannala (2015) argues that some models are better than others and, due to possible failures of implementation of these models, suggests the use of two or more programs based on the same or very similar statistical models.

In parallel with the developments in genetic species delimitation, there has been a renewed push to seek congruence across multiple lines of evidence – "integrative taxonomy" in order to arrive at more reliable and stable species delimitations. This reflects concern that traditional morphology-based methods struggle to identify cryptic species or can be misled due to convergent evolution of traits (Padial et al. 2010; Fujita et al. 2012). Some approaches have been able to simultaneously include genetic, phenotypic and geographic information (Guillot et al. 2005; Solís-Lemus et al. 2015). Additionally, it was also suggested that species description should include diagnostic traits that go beyond the limitations by conserved morphological evidence, using for example DNA-based diagnosis (Renner 2016).

Ultimately these advancements in detection and description of species will benefit biodiversity assessments that are important for management and conservation, with increasing efforts to also consider phylogenetic diversity (e.g. Rosauer et al. 2016, 2017; González-Orozco et al. 2016).

1.3 Hybridization

1.3.1 Role of gene flow and introgression in species evolutionary histories

Gene flow is one of the forces affecting evolutionary change and the development of isolation-with-migration models has enabled biologists to study the role of gene flow during species divergence (Pinho & Hey 2010). In contrast with a strict isolation or allopatric divergence scenario, alternative models that include gene flow are isolation with migration with constant rates of migration since split, isolation after migration where populations initially diverge with gene flow which later ceases and divergence continues in isolation, and a third model in which populations diverge isolated with no gene flow for a period of time after which gene flow is initiated when populations come into secondary contact (reviewed in Sousa & Hey 2013).

Although largely studied at intraspecific level, gene flow between genetically distinguishable taxa (species or diverging populations), i.e. hybridization with production of viable hybrids, has mostly been under appreciated as an important process in speciation of vertebrates (Rheindt & Edwards 2011; Mallet et al. 2016), despite this being long recognized for plants (Anderson & Stebbins Jr. 1954; Grant 1981; Arnold 1997). The increased availability

of genetic markers or even whole genome analyses, have been suggesting that speciation with gene flow could be more common than previously thought (Nosil 2008) and has been detected in well-studied species like cichlids, *Heliconius* butterflies and Darwin's finches (Martin et al. 2013; Lamichhaney et al. 2015; Meier et al. 2017b).

The study of hybridization is relevant to understand how differentiated lineages respond when coming into secondary contact, particularly after a period of time where they have been evolving separately but reproductive isolation (RI) is not yet complete. This is observable in hybrid zones (e.g. Singhal & Moritz 2012); where parapatric lineages are in contact and depending on strength of RI, current hybridization might be frequent and an equilibrium between migration of alleles and selection against determines the width of a stable contact zone (reviewed in Barton & Hewitt 1985). When this equilibrium is disrupted, hybridization may lead to introgression (or "introgressive hybridization"), i.e. the incorporation through backcrossing of alleles from one lineage into the gene pool of a second, divergent lineage (Anderson & Hubricht 1938; Anderson 1949). Perhaps more frequently, hybridisation between closely related taxa can lead to collapse of species boundaries (Seehausen 2004).

Introgression among even phylogenetically divergent species can be observable as discordance among gene trees above that expected from mutational variance or incomplete lineage sorting (ILS; Mallet et al. 2016; Edwards et al. 2016). Depending on the taxon and scale of analysis, reticulate evolution can also be due to lateral gene transfer, especially in prokaryotes (Mallet et al. 2016).

Hybridization may have several distinct outcomes on speciation, at different stages and in different spatial contexts (reviewed in Abbott et al. 2013). On one hand, it can slow or reverse differentiation by allowing gene flow among diverging lineages; on the other hand, it may accelerate speciation via introgression of a few loci promoting adaptive divergence (Mallet 2007), such as the mimicry genes in *Heliconius* butterflies (The *Heliconius* Genome Consortium 2012) and a gene related to beak traits in Darwin's finches (Lamichhaney et al. 2015). The temporal context of hybridization consists of either one or more occurrences of secondary contact after a period of independent evolution or continuous contact with divergent selection. The spatial context may reflect habitat disturbance or range expansion (e.g. Abbott et al. 2003; Zinner et al. 2009), and hybridization may occur in hybrid zones or in complex habitat mosaics that arise due to local adaptation in an heterogeneous environment (reviewed in Harrison & Larson 2014).

1.3.2 Hybridization and environmental change

One reason why hybridization is an important research topic is due to the impacts of anthropogenic climate change, which allow for species to shift their range forming new hybrid zones, but with unknown consequence for species stability due to the fast rate of change (Allendorf et al. 2001; Chunco 2014). Understanding the dynamics of stable hybrid zones

with habitat heterogeneity and the adaptation of hybrid phenotypes to new environments is providing important theory for this field (Taylor et al. 2015).

Where reproductive isolation depends in some way on ecological specialization, a loss of environmental heterogeneity can increase hybridization between populations, that have diverged though adaptation to different environments, perhaps even collapsing them into one taxon (Seehausen et al. 2008). Regardless of the spatial scale, the decrease of landscape heterogeneity may increase the probability of hybridization, by relaxing ecological divergent selection among species and by making costs of mate choice too high in declining populations. In the case of a lasting environmental disturbance that triggered the decrease of landscape heterogeneity, a possible outcome can be the a reduction of biodiversity (Seehausen et al. 2008; Gilman & Behm 2011). Examples of recent anthropogenic disturbance causing speciation reversal are in the whitefish (Vonlanthen et al. 2012), stickleback fishes (Taylor et al. 2006), and in tree finches (Kleindorfer et al. 2014).

On the other hand, hybrid speciation theory suggests that with increasing environmental heterogeneity, new species can be generated by hybridization when new niches are available, ecological selection is strong with high hybrid fertility, and these are spatially well separated habitats of parental and hybrid species (Rieseberg 1997; Buerkle et al. 2000). Hybridization can be advantageous in new or perturbed habitats, as hybrid genotypes can be fitter in different habitats than their parental species (Dowling & Secor 1997; Burke & Arnold 2001; Arnold & Martin 2010), which has been well studied in sunflowers (Karrenberg et al. 2006), Irises (Taylor et al. 2009), and also in Darwin's finches (Grant & Grant 1996; Lamichhaney et al. 2018). Additionally, habitat change may also cause "hybridization of the environment", increasing contact of ecologically separated species which facilitates hybridization and creates novel habitats in which hybrids thrive (Anderson 1948; Anderson & Stebbins Jr. 1954). Hybridization has also been shown to instigate adaptive radiations, due to environmental changes that increase rates of hybridization and responses to selection, driving populations to diversify under divergent selection (Seehausen 2004; Meier et al. 2017a).

A large proportion of global vertebrate species diversity that arose by ecological speciation during the Quaternary (Seehausen et al. 2008) is predicted to be susceptible to hybridization-driven diversity change, considering that intrinsic postzygotic isolation may take 2-5 million years to complete (Coyne & Orr 2004) and the occurrence of postglacial range expansions enabling species coming into secondary contact. In fact, there is an ample body of research showing hybridization across multiple groups that changed their range during the climatic oscillations of the Quaternary (reviewed in Hewitt 2011).

More research into hybridization through secondary contact of allopatric species has been done, while little attention has been given to hybridization between secondarily sympatric species (for examples; see Bettles et al. 2005; Hasselman et al. 2014; Rohde et al. 2015). It is often assumed sympatric species evolved reproductive barriers to gene flow, but they might differ in microhabitat preferences or other life history trait occurring in allotopy (Johanet et al. 2011). Nonetheless, and following the logic of Seehausen et al. (2008), change

or decline of microhabitat can break ecological barriers and also promote hybridization between proximal species.

1.3.3 Approaches to detecting gene flow

Since the development of isolation-with-migration models, it has been possible to detect divergence between taxa in the presence of gene flow (reviewed in Sousa & Hey 2013). Initially based on sequence data from multiple loci, more recent approaches have used SNP data and the site frequency spectrum (SFS), which can allow for comparisons of models of strict isolation vs gene flow between two or three populations. Such approaches are based on a composite likelihood that approximates the joint allele frequency distribution between populations to estimate migration rate, e.g., the diffusion approximation ∂a∂i (diffusion approximations for demographic inference) method (Gutenkunst et al. 2009), or the more flexible simulation approach using fastsimcoal2 (fast sequential markov coalescent, Excoffier et al. 2013). Approximate Bayesian Computation (ABC, Shafer et al. 2015) can provide higher precision to estimated parameters, calculating posterior distributions by comparing simulated against observed data from thousands of loci, but it is much more computationally expensive. However, these model-based approaches can only infer which of the tested models is closer to the data, often provide poor estimates of gene flow timing and can struggle to separate models with continuous vs episodic gene flow (Roux et al. 2013).

Several statistical approaches also exist to infer from genetic data if gene flow is currently occurring between diverged lineages in secondary contact. The most commonly used approach is to identify recent gene flow with genotype-based population clustering methods such as the software Structure (Pritchard et al. 2000) or Admixture (Alexander et al. 2009), or to discriminate among types of recent hybrids (NewHybrids, Anderson & Thompson 2002). However, to then infer reliably rates of introgression where different lineages are in contact, clinal analysis of allelic frequencies across hybrid zones is more informative (Gompert & Alex 2011).

An approach that can efficiently infer past introgression between non-sister species while accounting for ILS is the ABBA-BABA test (Green et al. 2010). This is becoming the most widely used method with the capability to detect introgression across genomic regions in whole-genome studies (Martin et al. 2015). This test is based on counts of ancestral (A) and derived (B) alleles from biallelic SNPs from species in a pectinate four-taxon topology (includes one outgroup). The test uses two patterns of interest, ABBA and BABA, that correspond to incongruences relative to the expected profile, BBAA, due to either introgression or to ILS. With ILS alone these patterns would be equally frequent and Patterson's D statistic infers introgression from a significant excess of one pattern over the other. Several methods have been developed based on this test: based on allelic frequencies (Durand et al. 2011), Partitioned D-statistics (Eaton & Ree 2013) and five-taxon ABBA-BABA (Pease & Hahn 2015).

A phylogenetic approach to distinguish patterns of ILS from gene flow is to model phylogenetic networks where introgression is presented by nodes connecting hybridizing species (or ancestral braches) in a phylogenetic tree. The two main programs for estimating phylogenetic networks with the capacity to infer magnitude and direction of gene flow are PhyloNet (Than et al. 2008) and PhyloNetworks (Solís-Lemus et al. 2017). PhyloNet is computationally intensive, based on rooted triplets and includes methods of maximum likelihood (Yu et al. 2014), pseudolikelihood (Yu & Nakhleh 2015), and Bayesian inference from gene trees (Wen et al. 2016), biallelic markers (Zhu et al. 2018) and from sequences (multispecies network coalescent method, Wen & Nakhleh 2017). PhyloNetworks uses a maximum pseudolikelihood estimation of species network (SNaQ, (Solís-Lemus & Ané 2016) which is faster since it is based on concordance factors of unrooted quartets. Additionally, another recent sequenced-based multispecies network coalescent method has been developed within BEAST2 (Bouckaert et al. 2014) - SpeciesNetwork (Zhang et al. 2017). This approach is very computationally intensive but can be used to study hybrid speciation and introgression between sister and non-sister taxa.

1.4 The Australian Monsoonal Tropics Rainbow skinks

1.4.1 *Carlia* – The Rainbow Skinks

Because of their diverse ecology and morphology, and low dispersal, lizards are great set of model organisms for studying phylogeography, speciation (Camargo et al. 2010) and responses to climate change (Huey et al. 2009). In particular, Australian lizards represent a very species rich (>1000 species) and interesting group (Powney et al. 2010; Wilson & Swan 2017). Recent phylogenetic and macroevolutionary analyses have yielded new insights into the evolutionary origins and subsequent radiations of this large and high endemic fauna (Rabosky et al. 2014; Oliver & Hugall 2017). The skinks (Scincidae) are the largest family of lizards with 437 species just in Australia (Wilson & Swan 2017). In this continent, the skinks are represented with three clades that separately colonized a depauperated Australian fauna after the separation from Antarctica: the Sphenomorphus (25 Ma), Egernia (18 Ma) and the Eugongylus group (20 Ma) (Skinner et al. 2011).

The rainbow skinks are one of the better studied group of Australian lizards that belong to the Eugongylus group (Greer 1979), and these likely went through a rapid diversification during the mid-late Miocene (Dolman & Hugall 2008). The rainbow skinks consist in the genus *Carlia* Gray (1845) and this informal name is also associated with the *Lygisaurus* De Vis (1884) and *Liburnascincus* Wells & Wellington (1984) genera, and in total these three genera contain more than 60 species across Australia, New Guinea and Wallacea (from reptile-database.reptarium.cz). The molecular systematics of the Australian species in this clade have been studied with mtDNA phylogenies (Stuart-Fox et al. 2002; Couper et al. 2005), mtDNA and nuclear markers (Dolman & Hugall 2008) and further with morphological and karyotype data (Zug 2004; Donnellan et al. 2009). However, there are multiple poorly resolved relation-

ships or topological conflicts within and between the three genera across the different datasets. Even the more complete and better resolved phylogenetic tree in Dolman & Hugall (2008) has low support for several nodes. The species across *Carlia* and *Lygisaurus* in particular have low interspecific morphological variation (Storr 1974), hindering identification of taxonomic diagnostic traits and species delimitation based only on morphology.

Lygisaurus (litter-skinks) is a genus resurrected from synonymy with *Carlia* in Ingram & Covacevich (1989), sunk by Stuart-Fox et al. (2001) and resurrected again by Dolman & Hugall (2008), and which contains 11 Australian species. These are smaller and mostly lack the male breeding colours and scale keeling present in *Carlia*. This genus can be further distinguished from the latter by the number of premaxillary teeth and supradigital scales on the fourth toe (Ingram & Covacevich 1989). These species occur across rainforest, woodland and rocky habitats in the northeast of Australia and Papua New Guinea (Ingram & Covacevich 1989). *Liburnascincus* (four-fingered rock skinks) consists of four saxicoline species that are restricted to rocky environments (Wilson & Swan 2017) in north-eastern Australia. This genus may be distinguished from *Carlia* by a combination of several traits related to the scales, ear opening, size and coloration (Cogger 2014).

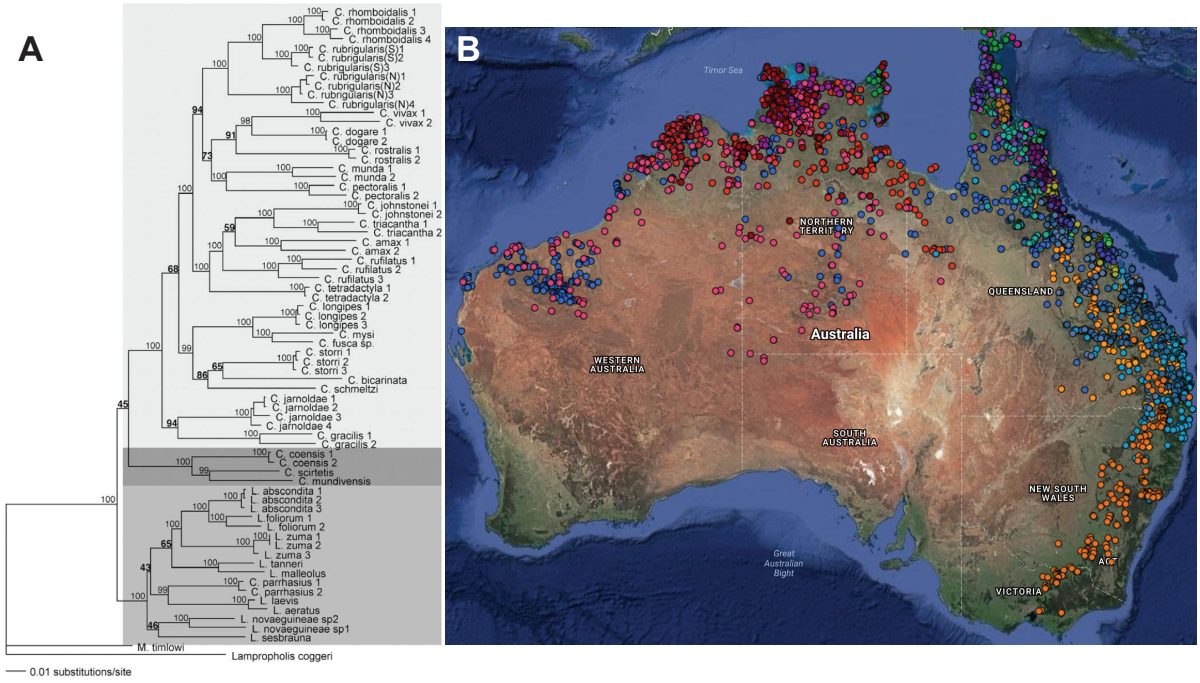


Figure 1.1 Phylogenetic relationships of the rainbow skinks from Dolman and Hugall (2008) (A), and distribution of *Carlia* species across Australia (B). Map and location points retrieved from Atlas of Living Australia website at www.ala.org.au. Accessed 4 February 2018.

Most species diversity (44 species overall, from reptile-database.reptarium.cz) is within *Carlia* (four-fingered skinks) with at least 24 species in Australia (Hoskin 2014) and many more across New Guinea and Wallacea, which are the focus of an ongoing taxonomic revision based on new phylogenomic data (E. Rittmeyer, in prep.). The Australian species are distributed across eastern Australia with a group of six species also occurring across the Australian

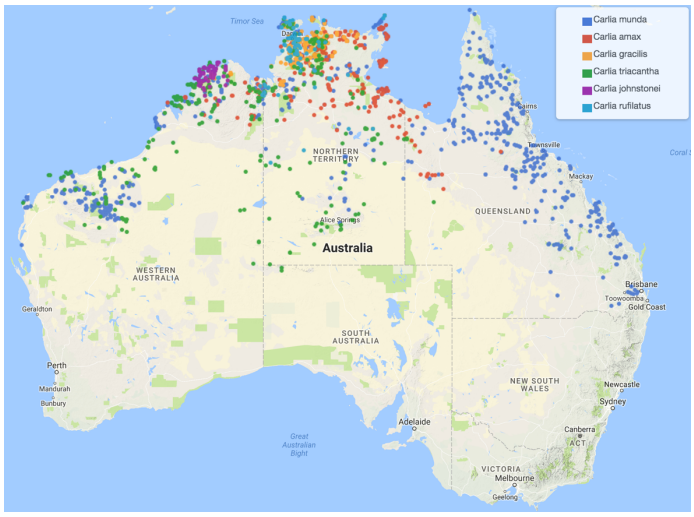


Figure 1.2 Distribution of the six Australian Monsoonal Tropics *Carlia* species. Map and location points retrieved from Atlas of Living Australia website at www.ala.org.au.

Monsoonal Tropics (AMT). Most species are in tropical and subtropical regions except for *C. munda* and *C. triacantha* that also extend to the arid centre, and *C. tetradactyla* in the more temperate southeast of the continent.

These skinks can be found in leaf litter and rocky areas across diverse vegetation types: tropical rainforest, spinifex grasslands, tropical woodlands, open eucalypt forest, agricultural lands and suburban gardens (Stuart-Fox et al. 2002); as well as diverse substrate types, from sandy soils in coastal and semi-arid areas to bare granite boulders (Ingram & Covacevich 1989; Cogger 2014). The highest abundance and species diversity is in dry woodland habitats and the rainforest from north-eastern Australia (Wilson & Swan 2017), where most research on this genus has been done. One main feature that characterizes these lizards is sexual dichromatism, in which males of most species develop bright lateral colouring during their wet-season mating period, whereas females lack bright colours (Ingram & Covacevich 1989). These small species (35 to 65 mm snout-vent length) are insectivorous, oviparous, actively foraging, diurnal, terrestrial skinks with well-developed limbs. Species identification can be achieved based on coloration, especially in the breeding season, along with presence, strength and number of dorsal keels, the relative size of transparent disc, the shape and alignment of ear-opening and ear-lobules characteristics (Wilson & Swan 2017).

Carlia has been the target of several recent taxonomical revisions (e.g. Zug 2004; Donnellan et al. 2009) and additional delimitations of species based on morphology and mtDNA data (Hoskin & Couper 2012; Hoskin 2014). Other published research concerns behaviour (Langkilde et al. 2003, 2005), ecomorphological studies (Goodman 2007; Dolman & Stuart-Fox 2010; Goodman et al. 2013; Manicom et al. 2014), sensitivity to fire (Braithwaite 1987) and physiological ecology (Singh et al. 2002; Vickers et al. 2011; Muñoz et al. 2016; Pintor et al. 2016a; b). These detailed studies are largely focused on species from north-east Australia. In parallel, there has been a lot of recent interest on speciation, responses to past climatic change and refugia with two rainforest-associated species, *C. rubrigularis* and C. *rhomboidalis* (Stuart-Fox et al. 2001; Phillips et al. 2004; Dolman & Moritz 2006; Dolman 2008; Moritz et al. 2012; Singhal & Moritz 2013; Singhal & Bi 2017).

By comparison with east Australian species of *Carlia*, much less research has been done on the Australian *Carlia* species that occur in the central and west of the AMT, long known as a biogeographically distinct region (Cracraft 1986; Edwards et al. 2016). This set of six

species - *C. johnstonei*, *C. triacantha*, *C. amax*, *C. rufilatus*, *C. munda* and *C. gracilis* - overlap in distribution and appear as a polyphyletic group of species in the phylogeny from Dolman & Hugall (2008, Fig. 1.1).

Figure 1.3 *Carlia johnstonei*

C. johnstonei and *C. triacantha* are each other's closest relatives and they overlap in distribution, however they have distinct climatic niches. *C. johnstonei* Storr, 1974 (rough brown rainbow skink) is the only endemic *Carlia* to the wettest region in the Kimberley (region in the north-west of Australia), and it is associated with monsoon forests of basalt country and leaf litter of other kinds of closed or dense vegetation (Ingram & Covacevich 1989). *C. triacantha* Mitchell, 1953 (desert rainbow skink) has a much wider distribution from the mesic north (Kimberley and Top End) to the arid centre of Australia (Pilbara and Central MacDonnell Ranges), and it is often found in sandstone spinifex, among other rocks, on sand and termite mounts (Ingram & Covacevich 1989). Morphologically these are more distinct than other pairs of sister species, with differences in number of keels in scales (two keels vs three), body coloration even without breeding colours, distance in prefrontal scales (usually separated vs in contact) and number of lobules inside ear-opening (many vs a few) (Cogger 2014).

Figure 1.4 *Carlia triacantha*

Figure 1.5 *Carlia amax*

Sister to this group is *Carlia* amax Storr, 1974 (two-spined or bauxite rainbow skink), with differences to *C. johnstonei* in body size, coloration and ear lobules (one anterior), similar with two keels on body dorsal scales, but with a broader distribution across the AMT. This is a fire-sensitive species (Nicholson et al. 2006) found mainly on rocks (especially laterite), and in spinifex or leaf litter of deciduous vine thickets growing at the foot of limestone outcrops and sandstone cliffs (Ingram & Covacevich 1989).

Carlia rufilatus Storr, 1974 (red-sided rainbow skink) is an outgroup to the previous three species, with a scattered distribution ranging from the Kimberley to the north-west of Top End where it is more common, and found in riparian rainforests, moderately dense forests and woodlands (Ingram & Covacevich 1989).

Figure 1.4 *Carlia rufilatus*

Figure 1.7 *Carlia munda*

This species has several morphological similarities with the distantly related *C. munda* De Vis, 1985 (striped or shaded rainbow skink): with a dotted or flecked with black coloration, white strip below eye to ear or forelimb, red lateral coloration in breeding males, prefrontals in contact and one to two ear lobules (Cogger 2014). These species differ in keeling strength; the latter is smooth to feebly tricarinate, whereas *C. rufilatus* is moderately keeled. These two species also differ in breadth of distribution and climates occupied (Potter et al. 2017); *C. munda* has a wide distribution, similar to *C. triacantha* with addition of an Australian East Coast distribution, whereas *C. rufilatus* is restricted to mesic areas of the Kimberley and Top End.

Figure 1.6 *Carlia gracilis*

The sixth species, *C. gracilis* Storr, 1974 (slender rainbow skink), is one of the more ancestral species within *Carlia* and it also occurs across the Kimberley and the Top End region. This species has a slender body shape with three moderately strong keels on dorsal scales, usually five supraciliaries, separated prefrontals (Cogger 2014), is fire sensitive and has a preference for abundant leaf litter of monsoon forests, coastal thickets and wetter woodland areas (Ingram & Covacevich 1989).

These last four species have been studied by Potter et al. (2016, 2017) to investigate lineage diversity and explore how concordant the species responses to late Pleistocene change were, finding cryptic lineages and dissimilar species responses between two regions. These studies found deep divergent lineages in islands of northern Australia, recent admixture and potential gene flow between some mainland intraspecific lineages, and detected in the Top End a more synchronous demographic expansion of lineages and climatic stability since the LGM than in comparison with the Kimberley region. It was also concluded that in areas of greater climate fluctuation and aridity prone like the Kimberley, species responses are more spatially and temporally variable reflecting differences in niche breadth.

Though relatively poorly studied, there are some ecological differences (e.g. fire sensitivity, timing of reproduction, habitat use) between these six species, but various combinations often occur in sympatry (James & Shine 1985; Braithwaite 1987; Gambold & Woinarski 1993; Dolman & Stuart-Fox 2010). They also have similar diets, behaviours and breeding seasons (Brown 1984; James & Shine 1985). Despite differences in range size and associated climate space, these species broadly overlap across several regions within the AMT, particularly in the mesic zones of the north-west Top End and the west Kimberley.

1.4.2 Biogeography of the Australian Monsoonal Tropics

The Australian Monsoonal tropics (Fig. 1.9) is a sparsely populated region with one of the largest, ecologically dynamic and pristine tropical savannas, and an underestimated high biodiversity value (Woinarski et al. 2007; Bowman et al. 2010). This region in the north of Australia is characterized by a latitudinal gradient of high to moderate wet summer monsoon (more than 85% rainfall between December to April) followed by a longer dry season during the rest of the year (Suppiah 1992; Bowman et al. 2010). The distribution of different habitats within the savanna (e.g. sandstone uplands, floodplains, mesic forests and freshwater wetlands) comes from the interaction of this monsoonal climate with topography, hydrology and fire behaviour (Woinarski et al. 2007; Bowman et al. 2010).

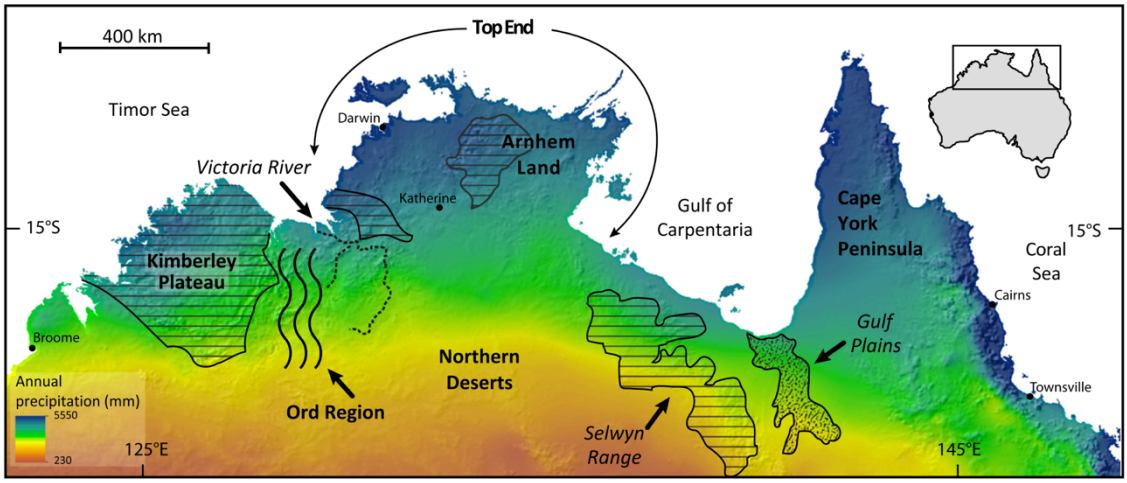


Figure 1.8 The Australian Monsoonal Tropics (AMT) mapped with main biogeographic regions and coloured by annual precipitation. The major areas shown with horizontal hatching represent raised or dissected terrain. Map from Moritz *et al.* (2016).

This biome is relatively understudied in comparison with the older and narrowly distributed rainforests of the Australian Wet Tropics (Williams et al. 1996; Schneider et al. 1998; Yeates et al. 2002; Pearson et al. 2015) or the younger and widespread arid zone (reviewed in Byrne et al. 2008; Oliver & Hugall 2017), both of which flank the AMT. The warm, wet climate and mesic biome that covered most of Australia started to decline with the aridification across most of central Australia in the mid-Miocene (Martin 2006; Fujioka & Chappell 2010). Around 5.5 million years ago (MYA) northwest Australia experienced a rapid transition from a dry to humid climate, with year-round rainfall during the Pliocene, until 3.3 MYA (Christensen et al. 2017). The climate then shifted to an arid with seasonal precipitation regime in the early Pleistocene around 2.4 MYA (Christensen et al. 2017). However, paleoecological evidence suggests that the direction and magnitude of climatic fluctuations from the last glacial maximum (LGM) to the present varied considerably across northern Australia (Reeves et al. 2013).

The biogeography of the AMT was summarized by Bowman et al. (2010) who summarized evidence for several barriers based on a combination of climatic and topological

factors (further revised in Eldridge et al. 2011; Catullo et al. 2014; Edwards et al. 2017). The three main biogeographic regions identified within the AMT (Fig. 1.9) are, from east to west: Cape York (Queensland), Top End (Northern Territory) and the Kimberley (Western Territory). The Cape York region is delineated by the Carpentarian Gap to the west and the Burdekin Gap to the south (reviewed in Chapple et al. 2011). The Top End is isolated to the west by either of both of the Victoria or Daly River catchments and to the south by the open woodlands of the Northern Deserts (Catullo et al. 2014). This region is ecologically and topographically diverse with high levels of biodiversity and endemism (Bowman et al. 2010; Rosauer et al. 2016).

The Kimberley region corresponds to the uplifted Kimberley plateau, the uplands of the King Leopold and Durack ranges (Li 2000; Pepper & Keogh 2014). It is isolated to the south by sandplains and increasing aridity and east by the Ord – Victoria River region. Biogeographical breaks within the Kimberley have been proposed by Pepper & Keogh (2014) and some recent studies of vertebrates support some parts of the region also as a Pleistocene refugium (e.g. Pepper et al. 2011a; Potter et al. 2012; Williams et al. 2013; Oliver et al. 2014; Catullo et al. 2014; Laver et al. 2018). Further evidence of persistence of species in this region comes from the discovery of large endemic radiations of land snails (Köhler & Criscione 2015). This notion of evolutionary refugia in the Kimberley contrasts with earlier views of complete displacement of monsoonal woodlands northwards to the exposed Arafura shelf during the LGM (Woinarski et al. 2007). Monsoon activity was weaker during the last glacial maximum (LGM) which likely resulted in dry and cool conditions until the early Holocene, changing around 14,000 years ago (Field et al. 2017) to wetter and warmer conditions than at present (Reeves et al. 2013; Proske et al. 2014) and then a gradual weakening of the monsoon activity.

The challenges for research across the several regions within the AMT are mostly due to remoteness, inaccessibility during wet season and restricted access regions located in Aboriginal-owned lands (Moritz et al. 2013). Despite this, there has been substantial recent effort, including extensive field work in association with Indigenous Rangers and other landholders, to increase the knowledge of biodiversity and richness across the region. These studies have been showing substantial, taxonomically unrecognised diversity with most species including several morphologically cryptic lineages, only some of which have subsequently been described. Examples of this rampant cryptic diversity include geckos (Oliver et al. 2010, 2013; Pepper et al. 2011b; Oliver & Parkin 2014; Moritz et al. 2016, 2017; Laver et al. 2018), frogs (Catullo et al. 2014), skinks (Potter et al. 2016, 2017), birds (Toon et al. 2010; Kearns et al. 2014), fishes (Huey et al. 2012), mammals (Potter et al. 2012, 2014, 2015; Wadley et al. 2016), and invertebrates (Braby 2008; Criscione et al. 2012; Andersen et al. 2014). For the increasingly well-studied reptiles, it is likely that the described species diversity of the Top End and Kimberley will increase several-fold.

These cryptic lineages are often within widespread taxonomic species and failure to recognize them obscures the evolutionary and biogeographic history of the region. Some

cryptic, often short-range, endemic lineages represent older events of vicariant differentiation, while widespread lineages often represent more recent events of dispersal and range expansions (Oliver et al. 2014; Pepper & Keogh 2014; Moritz et al. 2017). Furthermore, these recent studies highlight the necessity of extending high resolution phylogeographic analyses to other species in the region as such studies can provide spatial information that can be of valuable input for conservation and management of the rich ecosystems across the AMT (Moritz et al. 2013; Rosauer et al. 2016, 2018).

1.5 Aims and structure of this thesis

How speciation, range shifting and reticulation through climatic oscillations combine to shape current patterns of local and regional diversity remains a key question in evolution and ecology. Insights into these processes can be obtained through multi-scale analyses of closely related species in diverse communities, such as the rainbow skinks from the Australian Monsoonal Tropics (AMT). Hence, the main goal of this thesis is to study these species' evolutionary histories to better understand the processes that drive diversity across this rich and understudied Australian biome. In particular, I apply the custom exon-capture and sequencing approach (Bi et al. 2012; Bragg et al. 2016) to resolve demographic and speciation history, delimit species and infer their relationships.

More specifically the aims are to:

1. Identify lineage diversity within and across a pair of sympatric species;
2. Apply a combined genomic and morphological approach to validate new major lineages as separate species;
3. Compare the responses of species with starkly different observed climatic breadth to the same climatic fluctuations;
4. Infer the relationships of the AMT species to each other and other rainbow skinks using a phylogenomic approach;
5. Explore the potential for reticulation between these often sympatric AMT species, considering two regions with distinct climatic histories.

These five objectives were addressed across a total of six chapters. The present Chapter I, provided an overview of theoretical and methodological content, the study species and region, to introduce the necessary context to address these five aims.

Chapter II corresponds to a publication in the Molecular Ecology journal and addressed the first and third aim with the main focus to compare the responses to the most recent events of past climatic change of two sympatric sister species: the climatic specialist narrow endemic *C. johnstonei* with the widespread climatic generalist *C. triacantha*. To do this I used a broad geographic sampling and an exon capture dataset to firstly identify major intraspecific lineages with their divergence histories, and then to compare intraspecific patterns between overlapping lineages in the Kimberley region.

In Chapter III, the lineages that were identified in chapter II were further analysed, with the objective to test support for describing them as new species. I used an integrative approach with morphological data and multispecies coalescent methods, to statistically validate these lineages as independent units and then described them as new species, accomplishing the second aim and published in the journal PeerJ.

Chapter IV is a collaborative study published in BMC Evolutionary Biology journal, that addressed the fourth aim to infer the relationships of the *Carlia* species that broadly co-occur in the Australian Monsoonal Tropics. This is achieved by investigating the phylogenomics of the rainbow skinks, by using the same exon capture set to produce a well-resolved phylogeny of the Australian *Carlia* species.

Chapter V is a study in the final stage of preparation to submit to an indexed international scientific journal and it is focused on the last aim. Here, I explored reticulation across the six sympatric AMT species and their overlapping regional lineages identified in previous studies and Chapter II, using similar exon capture data and recent statistical methods. Specifically, I intended to compare introgression between overlapping lineages in the Kimberley vs Top End to identify the potential for a differential hybridization response in a more unstable vs stable region in the late Pleistocene.

Finally, the integration of the results in the different Chapters is provided in the General Discussion (Chapter VI). This discusses cryptic speciation and responses to past climatic change while using next-generation sequencing data, and includes future perspectives and final considerations.

1.6 References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Abbott R, James JK, Milne RI, Gillies ACM (2003) Plant introductions, hybridization and gene flow. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **358**, 1123 LP-1132.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, **19**, 1655–1664.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.

Andersen AN, Bocciarelli D, Fairman R, Radford IJ (2014) Conservation status of ants in an iconic region of monsoonal Australia: levels of endemism and responses to fire in the eastern Kimberley. *Journal of Insect Conservation*, **18**, 137–146.

Anderson E (1948) Hybridization of the Habitat. *Evolution*, **2**, 1–9.

Anderson E (1949) *Introgressive hybridization*. John Wiley & Sons, Inc., New York.

Anderson E, Hubricht L (1938) Hybridization in Tradescantia. III. The evidence for introgressive hybridization. *American Journal of Botany*, 396–402.

Anderson E, Stebbins Jr. GL (1954) Hybridization as an Evolutionary Stimulus. *Evolution*, **8**, 378–388.

Anderson EC, Thompson EA (2002) A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics*, **160**, 1217–1229.

Angilletta MJ (2009) Thermal Adaptation: *A Theoretical and Empirical Synthesis*. Oxford University Press, Oxford.

Arnold ML (1997) *Natural hybridization and evolution*. Oxford University Press on Demand.

Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. *Trends in Ecology & Evolution*, **25**, 530–536.

Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard university press.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Avise JC, Bowen BW, Ayala FJ (2016) In the light of evolution X: comparative phylogeography. *Proceedings of the National Academy of Sciences*, **113**, 7957–7961.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.

Bell RC, Parra JL, Tonione M *et al.* (2010) Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Molecular Ecology*, **19**, 2531–44.

Bettles CM, Docker MF, Dufour B, Heath DD (2005) Hybridization dynamics between sympatric species of trout: loss of reproductive isolation. *Journal of Evolutionary Biology*, **18**, 1220–1233.

Bi K, Vanderpool D, Singhal S *et al.* (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC genomics*, **13**, 403.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A software platform for bayesian evolutionary analysis. *PLOS Computational Biology*, **10**, e1003537.

Bowen BW, Gaither MR, DiBattista JD *et al.* (2016) Comparative phylogeography of the ocean planet. *Proceedings of the National Academy of Sciences*, **113**, 7962 LP-7969.

Bowman DMJS, Brown GK, Braby MF *et al.* (2010) Biogeography of the Australian monsoon tropics. *Journal of Biogeography*, **37**, 201–216.

Braby MF (2008) Biogeography of butterflies in the Australian monsoon tropics. *Australian Journal of Zoology*, **56**, 41.

Bragg JG, Potter S, Bi K *et al.* (2017) Resources for phylogenomic analyses of Australian terrestrial vertebrates. *Molecular Ecology Resources*, **17**, 869–876.

Bragg JG, Potter S, Bi K, Moritz C (2016) Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources*, **16**, 1059–1068.

Braithwaite RW (1987) Effects of fire regimes on lizards in the wet-dry tropics of Australia. *Journal of Tropical Ecology*, **3**, 265–275.

Brandley MC, Bragg JG, Singhal S *et al.* (2015) Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian Eugongylus group scincid lizards. *BMC Evolutionary Biology*, **15**, 62.

Brito PH, Edwards S V (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439–55.

Brown JH (1984) On the relationship between abundance and distribution of species. *The American Naturalist*, **124**, 255–279.

Brown JH (1995) *Macroecology*. University of Chicago Press.

Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of northwestern North America: a synthesis. *Special Publication-British Ecological Society*, **14**, 319–340.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity*, **84**, 441.

Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics*, **35**, 31–52.

Byrne M (2008) Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews*, **27**, 2576–2585.

Byrne M, Steane D a., Joseph L *et al.* (2011) Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography*, **38**, 1635–1656.

Byrne M, Yeates DK, Joseph L *et al.* (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology*, **17**, 4398–417.

Camargo A, Sinervo B, Sites JW (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology*, **19**, 3250–70.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Catullo RA, Lanfear R, Doughty P, Keogh JS (2014) The biogeographical boundaries of northern Australia: evidence from ecological niche models and a multi-locus phylogeny of *Uperoleia* toadlets (Anura: Myobatrachidae). *Journal of Biogeography*, **41**, 659–672.

Chapple DG, Hoskin CJ, Chapple SNJ, Thompson MB (2011) Phylogeographic divergence in the widespread delicate skink (*Lampropholis delicata*) corresponds to dry habitat barriers in eastern Australia. *BMC Evolutionary Biology,* **11**, 191.

Chejanovski Z., Wiens JJ (2014) Climatic niche breadth and species richness in temperate treefrogs. *Journal of Biogeography*, **41**, 1936–1946.

Christensen BA, Renema W, Henderiks J *et al.* (2017) Indonesian throughflow drove Australian climate from humid Pliocene to arid Pleistocene. *Geophysical Research Letters*, **44**, 6914–6925.

Chunco AJ (2014) Hybridization in a warmer world. *Ecology and Evolution*, **4**, 2019–2031.

Clark AG, Eisen MB, Smith DR *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, **450**, 203–218.

Cogger H (2014) *Reptiles and amphibians of Australia*. New Holland, Sydney.

Cook JA, Bidlack AL, Conroy CJ *et al.* (2001) A phylogeographic perspective on endemism in the Alexander Archipelago of southeast Alaska. *Biological Conservation*, **97**, 215–227.

Couper PJP, Wilmer JJW, Roberts L, Amey APA, Zug GGR (2005) Skinks currently assigned to *Carlia aerata* (Scincidae: Lygosominae) of north-eastern Queensland: a preliminary study of cryptic diversity and two new species. *Australian Journal of Zoology*, **53**, 35–49.

Coyne JA, Orr HA (2004) *Speciation.* Sinauer & Associates, Sunderland, Massachusetts.

Cracraft J (1986) Origin and evolution of continental biotas: speciation and historical congruence within the Australian avifauna. *Evolution*, **40**, 977–996.

Criscione F, Law ML, Köhler F (2012) Land snail diversity in the monsoon tropics of Northern Australia: revision of the genus *Exiligada* Iredale, 1939 (Mollusca: Pulmonata: Camaenidae), with description of 13 new species. *Zoological Journal of the Linnean Society*, **166**, 689–722.

Cutter AD (2013) Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution*. **69**(3), 1172-85

Davis M, Shaw R, Etterson J (2005) Evolutionary responses to changing climate. *Ecology*, **86**, 1704–1714.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–340.

Dolman G (2008) Evidence for differential assortative female preference in association with refugial isolation of rainbow skinks in Australia's tropical rainforests. *PloS One*, **3**, e3499.

Dolman G, Hugall AF (2008) Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Molecular Phylogenetics and Evolution*, **49**, 782–94.

Dolman G, Moritz C (2006) A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution*, **60**, 573–582.

Dolman G, Stuart-Fox D (2010) Processes driving male breeding colour and ecomorphological diversification in rainbow skinks: a phylogenetic comparative test. *Evolutionary Ecology*, **24**, 97–113.

Donnellan S, Couper P, Saint K, Wheaton L (2009) Systematics of the *Carlia* "*fusca*" complex (Reptilia: Scincidae) from northern Australia. *Zootaxa*, **2227**, 1–31.

Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*, **28**, 593–619.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using radseq data: an example from flowering plants (pedicularis: orobanchaceae). *Systematic Biology*, **62**, 689–706.

Edwards RD, Crisp MD, Cook DH, Cook LG (2017) Congruent biogeographical disjunctions at a continent-wide scale: quantifying and clarifying the role of biogeographic barriers in the Australian tropics. *PLoS One*, **12**, 1–17.

Edwards S V, Liu L, Pearl DK (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences*, **104**, 5936–5941.

Edwards S V, Potter S, Schmitt CJ, Bragg JG, Moritz C (2016) Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences*, **113**, 8025–8032.

Eldridge MDB, Potter S, Cooper SJB (2011) Biogeographic barriers in north-western Australia: an overview and standardisation of nomenclature. *Australian Journal of Zoology*, **59**, 270–272.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLOS One, **6**, e19379.

Espíndola A, Pellissier L, Maiorano L *et al.* (2012) Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. *Ecology Letters*, **15**, 649–657.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics,* **9**, e1003905.

Faircloth BC, McCormack JE, Crawford NG *et al.* (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717–726.

Favre A, Päckert M, Pauls SU *et al.* (2015) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews*, **90**, 236–253.

Feliner GN (2011) Southern European glacial refugia: a tale of tales. *Taxon*, **60**, 365–372.

Field E, McGowan HA, Moss PT, Marx SK (2017) A late Quaternary record of monsoon variability in the northwest Kimberley, Australia. *Quaternary International*, **449**, 119–135.

Fujioka T, Chappell J (2010) History of Australian aridity: chronology in the evolution of arid landscapes. *Geological Society, London, Special Publications*, **346**, 121–139.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, **27**, 480–8.

Gambold N, Woinarski JCZDA-1993 (1993) Distributional patterns of herpetofanua in monsoon rainforests of Northern Territory, Australia. *Australian Journal of Zoology*, **18,** 431–449.

Gaston KJ (1994) *Rarity*. Chapman & Hall, London.

Ghalambor CK (2006) Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integrative and Comparative Biology*, **46**, 5–17.

Gilman RT, Behm JE (2011) Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. *Evolution*, **65**, 2592–2605.

Gompert Z, Alex BC (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

González-Orozco CE, Pollock LJ, Thornhill AH *et al.* (2016) Phylogenetic approaches reveal biodiversity threats under climate change. *Nature Climate Change*, **6**, 1110–1114.

Goodman BA (2007) Divergent morphologies, performance, and escape behaviour in two tropical rock-using lizards (Reptilia: Scincidae). *Biological Journal of the Linnean Society*, **91**, 85–98.

Goodman B a, Schwarzkopf L, Krockenberger AK (2013) Phenotypic integration in response to incubation environment adaptively influences habitat choice in a tropical lizard. *The American Naturalist*, **182**, 666–73.

Grant V (1981) *The Syngameon*. In: Plant Speciation, pp. 235–241. Columbia University Press, New York.

Grant BR, Grant PR (1996) High survival of Darwin's finch hybrids: effects of beak morphology and diets. *Ecology*, **77**, 500–509.

Green RE, Krause J, Briggs AW *et al.* (2010) A Draft Sequence of the Neanderthal Genome. *Science*, **328**, 710–722.

Greer AE (1979) A phylogenetic subdivision of Australian skinks. *Records of the Australian Museum*, **32**, 339–371.

Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, **5**, e1000695.

Hardy OJ, Born C, Budde K *et al.* (2013) Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience*, **345**, 284–296.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795–809.

Hasselman DJ, Argo EE, McBride MC *et al.* (2014) Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Molecular Ecology*, **23**, 1137–1152.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*, **10**, 537–49.

Hewitt GM (2011) Quaternary phylogeography: The roots of hybrid zones. *Genetica*, **139**, 617–638.

Hickerson MJ, Carstens BC, Cavender-Bares J *et al.* (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.

Hoskin CJ (2014) A new skink (Scincidae: *Carlia*) from the rainforest uplands of Cape Melville, northeast *Australia. Zootaxa*, **3869**, 224–236.

Hoskin CJ, Couper PJ (2012) Description of two new *Carlia* species (Reptilia: Scincidae) from northeast Australia, elevation of *Carlia pectoralis inconnexa* (Ingram & Covacevich 1989) to full species status, and redescription of *Carlia pectoralis* (de Vis 1884). *Zootaxa*, **28**, 1–28.

Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: bayesian inference of population structure. *Evolutionary Bioinformatics Online*, **7**, 55–59.

Huey RB, Deutsch CA, Tewksbury JJ *et al.* (2009) Why tropical forest lizards are vulnerable to climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 1939–48.

Huey RB, Kearney MR, Krockenberger A *et al.* (2012) Predicting organismal vulnerability to climate warming: roles of behaviour, physiology and adaptation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 1665–79.

Huey RB, Kingsolver JG (1993) Evolution of resistance to high temperature in ectotherms. *The American Naturalist,* **142**, S21-S46.

Hugall AF, O'hara TD, Hunjan S, Nilsen R, Moussalli A (2016) An exon-capture system for the entire class Ophiuroidea. *Molecular Biology and Evolution*, **33**, 281–294.

Ingram G, Covacevich JA (1989) Revision of the genus *Carlia* (Reptilia, Scincidae) in Australia and comments on *Carlia bicarinata* of New Guinea. *Memoirs of the Queensland Museum*, **27**, 443–490.

James C, Shine R (1985) The seasonal timing of reproduction: a tropical-temperate comparison in Australian lizards. *Oecologia*, **67**, 464–474.

Janzen DH (1967) Why mountain passes are higher in the tropics. *The American Naturalist*, **101**, 233–249.

Jarvis ED, Mirarab S, Aberer AJ *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Johanet A, Secondi J, Lemaire C (2011) Widespread introgression does not leak into allotopy in a broad sympatric zone. *Heredity*, **106**, 962–972.

Johnson CN (1998) Species extinction and the relationship between distribution and abundance. *Nature*, **394**, 272.

Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, **25**, 185–202.

Karrenberg S, Edelist C, Lexer C, Rieseberg L (2006) Response to salinity in the homoploid hybrid species *Helianthus paradoxus* and its progenitors *H. annuus* and *H. petiolaris*. *New Phytologist*, **170**, 615–629.

Kearns AM, Joseph L, Toon A, Cook LG, Fleischer RC (2014) Australia's arid-adapted butcherbirds experienced range expansions during Pleistocene glacial maxima. *Nature Communications*, **5**, 1153–1167.

Kleindorfer S, O'Connor JA, Dudaniec RY *et al.* (2014) Species collapse via hybridization in Darwin's tree finches. *The American Naturalist*, **183**, 325–341.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.

Köhler F, Criscione F (2015) A molecular phylogeny of camaenid land snails from north-western Australia unravels widespread homoplasy in morphological characters (Gastropoda, Helicoidea). *Molecular Phylogenetics and Evolution*, **83**, 44–55.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**, 17–24.

Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371.

Lamichhaney S, Han F, Webster MT *et al.* (2018) Rapid hybrid speciation in Darwin's finches. *Science*, **359**, 224 LP-228.

Langkilde T, Schwarzkopf L, Alford R (2003) An ethogram for adult male rainbow skinks, *Carlia jarnoldae*. *Herpetological journal*, **13**, 141–148.

Langkilde T, Schwarzkopf L, Alford RA (2005) The Function of Tail Displays in Male Rainbow Skinks (*Carlia jarnoldae*). *Journal of Herpetology*, **39**, 325–328.

Laver RJ, Doughty P, Oliver PM (2018) Origins and patterns of endemic diversity in two specialized lizard lineages from the Australian Monsoonal Tropics (*Oedura* spp.). *Journal of Biogeography*, **45**, 142–153.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species Delimitation using Genome-Wide SNP Data. *Systematic Biology*, **63**, 534–42.

Leaché AD, Oaks JR (2017) The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **48**, 69–84.

Lemmon AR, Emme SA, Lemmon EM (2012) Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. *Systematic Biology*, **61**, 727–744.

Lemmon EM, Lemmon AR (2013) High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Li ZX (2000) Palaeomagnetic evidence for unification of the North and West Australian cratons by ca. 1.7 Ga: new results from the Kimberley Basin of northwestern Australia. *Geophysical Journal International*, **142**, 173–180.

Liu L (2008) BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*, **24**, 2542–2543.

Liu L, Pearl DK (2007) Species Trees from Gene Trees: Reconstructing Bayesian Posterior Distributions of a Species Phylogeny Using Estimated Gene Tree Distributions. *Systematic Biology*, **56**, 504–514.

Liu L, Wu S, Yu L (2015a) Coalescent methods for estimating species trees from phylogenomic data. *Journal of Systematics and Evolution*, **53**, 380–390.

Liu L, Xi Z, Wu S, Davis CC, Edwards S V (2015b) Estimating phylogenetic trees from genome-scale data. *Annals of the New York Academy of Sciences*, **1360**, 36–53.

Liu L, Yu L, Edwards S V (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology*, **10**, 302.

Lohman DJ, de Bruyn M, Page T *et al.* (2011) Biogeography of the Indo-Australian Archipelago. *Annual Review of Ecology, Evolution, and Systematics*, **42,** 205–226.

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, **21**, 3656–3670.

Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–83.

Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays*, **38**, 140–149.

Manicom C, Alford R, Schoener TW, Schwarzkopf L (2014) Mechanisms causing variation in sexual size dimorphism in three sympatric, congeneric lizards. *Ecology*, **95**, 1531–44.

Martin HA (2006) Cenozoic climatic change and the development of the arid vegetation in Australia. *Journal of Arid Environments*, **66**, 533–563.

Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome research*, **23**, 1817–1828.

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, **32**, 244–257.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular phylogenetics and evolution*, **66**, 526–38.

Meier JI, Marques DA, Mwaiko S *et al.* (2017a) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, **8,** 14363.

Meier JI, Sousa VC, Marques DA *et al.* (2017b) Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Molecular ecology*, **26**, 123–141.

Mendes FK, Hahn MW (2016) Gene Tree Discordance Causes Apparent Substitution Rate Variation. *Systematic Biology*, **65**, 711–721.

Meng HH, Gao XY, Huang JF, Zhang ML (2015) Plant phylogeography in arid Northwest China: Retrospectives and perspectives. *Journal of Systematics and Evolution*, **53**, 33–46.

Mirarab S, Reaz R, Bayzid MS et al. (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics, **30**,* 541-8.

Moritz C, Agudo R (2013) The future of species under climate change: Resilience or decline? *Science*, **341**, 504–508.

Moritz C, Ens EJ, Potter S, Catullo RA (2013) The Australian monsoonal tropics: An opportunity to protect unique biodiversity and secure benefits for Aboriginal communities. *Pacific Conservation*, **19**, 343–355.

Moritz C, Fujita MK, Rosauer D *et al.* (2016) Multilocus phylogeography reveals nested endemism in a gecko across the monsoonal tropics of Australia. *Molecular Ecology*, **25**, 1354–1366.

Moritz C, Langham G, Kearney MR *et al.* (2012) Integrating phylogeography and physiology reveals divergence of thermal traits between central and peripheral lineages of tropical rainforest lizards. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 1680–1687.

Moritz CC, Pratt RC, Bank S *et al.* (2017) Cryptic lineage diversity, body size divergence, and sympatry in a species complex of Australian lizards (*Gehyra*). *Evolution*, **72**, 54–66.

Moussalli A, Moritz C, Williams SE, Carnaval AC (2009) Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. *Molecular Ecology*, **18**, 483–99.

Muñoz MM, Langham GM, Brandley MC *et al.* (2016) Basking behavior predicts the evolution of heat tolerance in Australian rainforest lizards. *Evolution*, **70**, 2537–2549.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.

Nicholson E, Lill A, Andersen A (2006) Do tropical savanna skink assemblages show a short-term response to low-intensity fire? *Wildlife Research*, **33**, 331–338.

Nieto Feliner G (2014) Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics,* **16**, 265–278.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59–73.

Ogilvie HA, Bouckaert R, Drummond AJ (2017) StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution*, **34**, 2101–2114.

Ogilvie HA, Heled J, Xie D, Drummond AJ (2016) Computational performance and statistical accuracy of *BEAST and comparisons with other methods. *Systematic Biology*, **65,** 381–396.

Oliver PM, Adams M, Doughty P (2010) Molecular evidence for ten species and Oligo-Miocene vicariance within a nominal Australian gecko species (*Crenadactylus ocellatus*, Diplodactylidae). *BMC Evolutionary Biology,* **10**, 386.

Oliver PM, Doughty P, Palmer R *et al.* (2013) Hidden biodiversity in rare northern Australian vertebrates: the case of the clawless geckos (*Crenadactylus*, Diplodactylidae) of the Kimberley. *Wildlife Research*, **39**, 429.

Oliver PM, Hugall AF (2017) Phylogenetic evidence for mid-Cenozoic turnover of a diverse continental biota. *Nature Ecology and Evolution*, **1,** 1896–1902.

Oliver PM, Laver RJ, Melville J, Doughty P (2014) A new species of Velvet Gecko (*Oedura*: Diplodactylidae) from the limestone ranges of the southern Kimberley, Western Australia. *Zootaxa*, **3873**, 49–61.

Oliver PM, Parkin T (2014) A new phasmid gecko (Squamata: Diplodactylidae: *Strophurus*) from the Arnhem Plateau: more new diversity in rare vertebrates from northern Australia. *Zootaxa*, **3878**, 37–48.

Padial J, Miralles A, Riva ID la, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 1–14.

Pearson RG, Connolly NM, Boyero L (2015) Ecology of streams in a biogeographic isolate—the Queensland Wet Tropics, Australia. *Freshwater Science*, **34**, 797–819.

Pease JB, Hahn MW (2015) Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic Biology*, **64**, 651–662.

Pepper M, Doughty P, Hutchinson MN, Keogh JS (2011a) Ancient drainages divide cryptic species in Australia's arid zone: morphological and multi-gene evidence for four new species of Beaked Geckos (*Rhynchoedura*). *Molecular Phylogenetics and Evolution*, **61**, 810–22.

Pepper M, Ho SYW, Fujita MK, Keogh JS (2011b) The genetic legacy of aridification: climate cycling fostered lizard diversification in Australian montane refugia and left low-lying deserts genetically depauperate. *Molecular Phylogenetics and Evolution*, **61**, 750–9.

Pepper M, Keogh JS (2014) Biogeography of the Kimberley, Western Australia: a review of landscape evolution and biotic response in an ancient refugium. *Journal of Biogeography*, **41**, 1443–1455.

Phillips BL, Baird SJE, Moritz C (2004) When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, **58**, 1536–1548.

Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.

Pintor AF V, Schwarzkopf L, Krockenberger AK (2016a) Hydroregulation in a tropical dry-skinned ectotherm. *Oecologia*, **182**, 925–931.

Pintor AF V, Schwarzkopf L, Krockenberger AK (2016b) Extensive acclimation in ectotherms conceals interspecific variation in thermal tolerance limits. *PLOS one*, **11**, e0150408.

Pons J, Barraclough TG, Gomez-Zurita J *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, **25**, 1367–1380.

Potter S, Close RL, Taggart DA, Cooper SJB, Eldridge MDB (2015) Taxonomy of rock-wallabies, *Petrogale* (Marsupialia: Macropodidae). IV. Multifaceted study of the *brachyotis* group identifies additional taxa. *Australian Journal of Zoology*, **62**, 401–414.

Potter S, Eldridge MDB, Taggart DA, Cooper SJB (2012) Multiple biogeographical barriers identified across the monsoon tropics of northern Australia: phylogeographic analysis of the *brachyotis* group of rock-wallabies. *Molecular Ecology*, **21**, 1–16.

Potter S, Rosauer D, Doody JS, Webb MJ, Eldridge MDB (2014) Persistence of a potentially rare mammalian genus (*Wyulda*) provides evidence for areas of evolutionary refugia within the Kimberley, Australia. *Conservation Genetics*, **15**, 1085–1094.

Potter S, Xue AT, Bragg JG et al. (2017) Pleistocene climatic changes drive diversification across a tropical savanna. *Molecular Ecology*. In Press. doi: 10.1111/mec.14441.

Powney GD, Grenyer R, Orme CDL, Owens IPF, Meiri S (2010) Hot, dry and different: Australian lizard richness is unlike that of mammals, amphibians and birds. *Global Ecology and Biogeography*, **19**, 386–396.

Prates I, Xue AT, Brown JL *et al.* (2016) Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences*, **113**, 7978–7985.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945 LP-959.

Proske U, Heslop D, Haberle S (2014) A Holocene record of coastal landscape dynamics in the eastern Kimberley region, Australia. *Journal of Quaternary Science*, **29**, 163–174.

Puritz JB, Matz M V., Toonen RJ *et al.* (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937–5942.

Qiu Y-X, Fu C-X, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution*, **59**, 225–244.

Rabosky DL, Donnellan SC, Grundler M, Lovette IJ (2014) Analysis and Visualization of Complex Macroevolutionary Dynamics: An Example from Australian Scincid Lizards. *Systematic Biology*, **63**, 610–627.

Rannala B (2015) The art and science of species delimitation. *Current Zoology*, **61**, 846–853.

Rannala B, Yang Z (2017) Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*, **66**, 823–842.

Reeves JM, Bostock HC, Ayliffe LK *et al.* (2013) Palaeoenvironmental change in tropical Australasia over the last 30,000 years – a synthesis by the OZ-INTIMATE group. *Quaternary Science Reviews*, **74**, 97–114.

Renner SS (2016) A return to Linnaeus's focus on diagnosis, not description: The use of DNA characters in the formal naming of species. *Systematic Biology*, **65**, 1085–1095.

Rheindt FE, Edwards S V. (2011) Genetic Introgression: An Integral but neglected component of speciation in birds. *The Auk*, **128**, 620–632.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual review of Ecology and Systematics*, **28**, 359–389.

Rissler LJ (2016) Union of phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences*, 2016, 201601073.

Rittmeyer EN, Austin CC (2012) The effects of sampling on delimiting species from multi-locus sequence data. *Molecular Phylogenetics and Evolution*, **65**, 451–463.

Rohde K, Hau Y, Weyer J, Hochkirch A (2015) Wide prevalence of hybridization in two sympatric grasshopper species may be shaped by their relative abundances. *BMC Evolutionary Biology*, **15**, 1–14.

Rosauer DF, Blom MPK, Bourke G *et al.* (2016) Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia. *Biological Conservation*, **204**, 83–93.

Rosauer DF, Byrne M, Blom MPK *et al.* (2018) Real-world conservation planning for evolutionary diversity in the Kimberley, Australia, sidesteps uncertain taxonomy. *Conservation Letters*, e12438.

Rosauer DF, Pollock LJ, Linke S, Jetz W (2017) Phylogenetically informed spatial planning is required to conserve the mammalian tree of life. *Proceedings of the Royal Society B: Biological Sciences*, **284**, 20170627.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, **30**, 1574–1587.

Rubin BER, Ree RH, Moreau CS (2012) Inferring Phylogenies from RAD Sequence Data. *PLOS One*, **7,** e33394.

Schneider C, Cunningham M, Moritz C (1998) Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Molecular Ecology*, **7**, 487–498.

Schneider C, Moritz C (1999) Rainforest refugia and Australia's Wet Tropics. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **266**, 191-196.

Seehausen O (2004) Hybridization and adaptive radiation. Trends in Ecology & Evolution, **19**, 198–207.

Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.

Shafer A, Cullingham CI, Cote SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589–4621.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW (2015) Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. *Molecular Ecology*, **24**, 328–345.

Sims GE, Jun S-R, Wu GA, Kim S-H (2009) Whole-genome phylogeny of mammals: Evolutionary information in genic and nongenic regions. *Proceedings of the National Academy of Sciences*, **106**, 17077-17082.

Singhal S, Bi K (2017) History cleans up messes: The impact of time in driving divergence and introgression in a tropical suture zone. *Evolution*, **71**, 1888–1899.

Singhal S, Grundler M, Colli G, Rabosky DL (2017) Squamate Conserved Loci (SqCL): a unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources*, **17**, e12-e24.

Singhal S, Moritz C (2012) Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard. *Evolution*, **66,** 1474–1489.

Singhal S, Moritz C (2013) Reproductive isolation between phylogeographic lineages scales with divergence. *Proceedings of the Royal Society of London Series B, Biological Sciences*, **280**.

Singh S, Smyth AK, Blomberg SP (2002) Thermal ecology and structural habitat use of two sympatric lizards (*Carlia vivax* and *Lygisaurus foliorum*) in subtropical Australia. *Austral Ecology*, **27**, 616–623.

Skinner A, Hugall AF, Hutchinson MN (2011) Lygosomine phylogeny and the origins of Australian scincid lizards. *Journal of Biogeography*, **38**, 1044–1058.

Slatyer RA, Hirst M, Sexton JP (2013) Niche breadth predicts geographical range size: a general ecological pattern. *Ecology Letters*, **16**, 1104–1114.

Solís-Lemus C, Ané C (2016) Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting. *PLOS Genetics*, **12**, e1005896.

Solís-Lemus C, Bastide P, Ané C (2017) PhyloNetworks: A Package for Phylogenetic Networks. *Molecular Biology and Evolution*, **34**, 3292–3298.

Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, **69**, 492–507.

Solís-Lemus C, Yang M, Ané C (2016) Inconsistency of Species Tree Methods under Gene Flow. *Systematic Biology*, **65**, 843–851.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404–414.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–3.

Storr GM (1974) The genus *Carlia* (Lacertilia, Scincidae) in Western Australia and Northern Territory. *Records of the Western Australian Museum*, 3**,** 151–165.

Stuart-Fox DM, Hugall AF, Moritz C (2002) A molecular phylogeny of rainbow skinks (Scincidae: *Carlia*): taxonomic and biogeographic implications. *Australian Journal of Zoology*, **50**, 39–51.

Stuart-Fox D, Schneider C, Moritz C, Couper PJ (2001) Comparative phylogeography of three rainforest-restricted lizards from mid-east Queensland. *Australian Journal of Zoology*, **49**, 119–127.

Suppiah R (1992) The Australian summer monsoon: a review. *Progress in Physical Geography*, **16**, 283–318.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Taylor SJ, Arnold M, Martin NH (2009) The genetic architecture of reproductive isolation in Louisiana irises: hybrid fitness in nature. *Evolution*, **63**, 2581–2594.

Taylor EB, Boughman JW, Groenenboom M *et al.* (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, **15**, 343–355.

Taylor SA, Larson EL, Harrison RG (2015) Hybrid zones: windows on climate change. *Trends in Ecology & Evolution*, **30**, 398–406.

Tewksbury JJ, Huey RB, Deutsch CA (2008) Putting the Heat on Tropical Animals. *Science,* **320**, 1296-1297.

Than C, Ruths D, Nakhleh L (2008) PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*, **9**, 322.

The *Heliconius* Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Thomé MTC, Carstens BC (2016) Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. *Proceedings of the National Academy of Sciences*, **113**, 8010–8017.

Thuiller W, Lavorel S, Araújo MB (2005) Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology and Biogeography*, **14**, 347–357.

Toon A, Hughes JM, Joseph L (2010) Multilocus analysis of honeyeaters (Aves: Meliphagidae) highlights spatio-temporal heterogeneity in the influence of biogeographic barriers in the Australian monsoonal zone. *Molecular Ecology*, **19**, 2980–2994.

Turchetto-Zolet AC, Pinheiro F, Salgueiro F, Palma-Silva C (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.

Vachaspati P, Warnow T (2015) ASTRID: Accurate Species TRees from Internode Distances. *BMC Genomics*, **16**, S3–S3.

Vickers M, Manicom C, Schwarzkopf L (2011) Extending the cost-benefit model of thermoregulation: high-temperature environments. *The American Naturalist*, **177**, 452–61.

Vonlanthen P, Bittner D, Hudson AG *et al.* (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, **482**, 357.

Wadley JJ, Fordham DA, Thomson VA, Ritchie EG, Austin JJ (2016) Phylogeography of the antilopine wallaroo (*Macropus antilopinus*) across tropical northern Australia. *Ecology and Evolution*, **6**, 8050–8061.

Walters RJ, Blanckenhorn WU, Berger D (2012) Forecasting extinction risk of ectotherms under climate warming: An evolutionary perspective. *Functional Ecology*, **26**, 1324–1338.

Warnow T (2015) Concatenation analyses in the presence of incomplete lineage sorting. PLoS Currents, **7**.

Weiss S, Ferrand N (2007) *Phylogeography of southern European refugia*. Springer.

Wen D, Nakhleh L (2017) Coestimating reticulate phylogenies and gene trees from multilocus sequence data. *Systematic Biology*, syx085.

Wen D, Yu Y, Nakhleh L (2016) Bayesian inference of reticulate phylogenies under the multispecies network coalescent. *PLoS Genetics*, **12**, 1–17.

Werneck FP (2011) The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.

Wiens JJ (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution*, **58**, 193.

Wiens JJ, Graham CH (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519–539.

Williams SE, Pearson RG, Walsh PJ (1996) Distributions and biodiversity of the terrestrial vertebrates of Australia's wet tropics: A review of current knowledge. *Pacific Conservation Biology*, **2**, 327–362.

Williams AN, Ulm S, Cook AR, Langley MC, Collard M (2013) Human refugia in Australia during the last glacial maximum and terminal Pleistocene: a geospatial analysis of the 25-12ka Australian archaeological record. *Journal of Archaeological Science*, **40**, 4612–4625.

Wilson S, Swan G (2017) *A complete guide to reptiles of Australia*. New Holland.

Woinarski J, Mackey B, Nix H, Traill B (2007) *The nature of northern Australias, ecological processes and future prospects*. ANU Press, Canberra Australia.

Yang Z, Rannala B (2014) Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. *Molecular Biology and Evolution*, **31**, 3125–3135.

Yeates DK, Bouchard P, Monteith GB (2002) Patterns and levels of endemism in the Australian Wet Tropics rainforest: evidence from flightless insects. *Invertebrate Systematics*, **16**, 605–619.

Yu Y, Dong J, Liu KJ, Nakhleh L (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences*, **111**, 16448–16453.

Yu Y, Nakhleh L (2015) A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics*, **16**, S10.

Zamudio KR, Bell RC, Mason NA (2016) Phenotypes in phylogeography: species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences*, **113**, 8041–8048.

Zhang C, Ogilvie HA, Drummond AJ, Stadler T (2017) Bayesian inference of species networks from multilocus sequence data. *Molecular Biology and Evolution*, msx307.

Zhang C, Rannala B, Yang Z (2014) Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic Biology*, **63**, 993–1004.

Zhu J, Wen D, Yu Y, Meudt HM, Nakhleh L (2018) Bayesian inference of phylogenetic networks from bi-allelic genetic markers. *PLoS Computational Biology*, **14**, e1005932.

Zinner D, Groeneveld LF, Keller C, Roos C (2009) Mitochondrial phylogeography of baboons (*Papio* spp.): Indication for introgressive hybridization? *BMC Evolutionary Biology*, **9**, 83.

Zug GR (2004) Systematics of the *Carlia* "*fusca*" lizards (Squamata: Scincidae) of New Guinea and nearby islands. Bishop Museum Bulletin (Zoology), **5**, 1–83.

Chapter II

TROPICAL SPECIALIST VERSUS CLIMATE GENERALIST: DIVERSIFICATION AND DEMOGRAPHIC HISTORY OF SISTER SPECIES OF *CARLIA* SKINKS FROM NORTHWESTERN AUSTRALIA

Ana C. Afonso Silva[1,2], Jason G. Bragg[1,3], Sally Potter[1], Carlos Fernandes[2], Maria Manuela Coelho[2] and Craig Moritz[1]

[1]Research School of Biology and Centre for Biodiversity Analysis, Australian National University, Acton, ACT 2601, Australia
[2]cE3c - Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, 1749-016, Lisboa, Portugal
[3]Royal Botanic Garden, Sydney, NSW 2000, Australia

The original work described in this chapter has been published and should be referenced as:

Afonso Silva AC., Bragg JG., Potter S., Fernandes C., Coelho MM., Moritz C. (2017). Tropical specialist vs. climate generalist: Diversification and demographic history of sister species of *Carlia* skinks from northwestern Australia. *Molecular Ecology* **26**,4045–4058.

Abstract

Species endemic to the tropical regions are expected to be vulnerable to future climate change due in part to their relatively narrow climatic niches. In addition, these species are more likely to have responded strongly to past climatic change, and this can be explored through phylogeographic analyses. To test the hypothesis that tropical specialists are more sensitive to climate change than climate generalists, we generated and analyze sequence data from mtDNA and ~2500 exons to compare scales of historical persistence and population fluctuation in two sister species of Australian rainbow skinks: the tropical specialist *Carlia johnstonei* and the climate generalist *C. triacantha*. We expect the tropical specialist species to have deeper and finer-scale phylogeographic structure and stronger demographic fluctuations relative to the closely related climate generalist species, which should have had more stable populations through periods of harsh climate in the late Quaternary. Within *C. johnstonei*, we find that populations from the northern Kimberley islands are highly divergent from mainland populations. In *C. triacantha*, one major clade occurs across the deserts and into the mesic Top End, and another occurs primarily in the Kimberley with scattered records eastwards. Where their ranges overlap in the Kimberley, both mitochondrial and nuclear DNA suggest stronger phylogeographic structure and range expansion within the tropical specialist, whereas the climate generalist has minimal structuring and no evidence of recent past range expansion. These results are consistent with the hypothesis that tropical specialists are more sensitive to past climatic change.

Keywords: Exon capture, population genomics, Kimberley, Rainbow skinks

2.1 Introduction

Past responses to climate change are reflected in spatial patterns of genetic diversity, so it can be informative to compare these patterns between (Rödder et al. 2013) and within species (McDonough et al. 2015). It is usually expected that specialist species with relatively narrow climatic tolerances will have persisted through the climatic changes of the late Pleistocene in one or a few geographic refugia and experienced strong demographic fluctuations that might still be detectable in patterns of genetic variation (e.g. Moussalli et al. 2009; Bell et al. 2010; Werneck et al. 2011; López-Uribe et al. 2014). By contrast, species with broader climatic niches are mostly predicted to be less affected, with smaller demographic or spatial change (e.g. Janzen 1967, Peres et al. 2015; Berkman et al. 2015).

Tropical ectotherms are assumed to have narrow thermal niches (Sunday et al. 2011; Huey et al. 2012), leading to the expectation that they will be more sensitive to climate change than widespread or generalist species. Accordingly, much attention has been given to testing for and understanding vulnerability of tropical taxa to climate change (Deutsch et al. 2008; Huey et al. 2009), with various studies focusing on geographic range size and niche breadth shifts (Pither 2003; McCain 2009; Pincheira-Donoso 2011; Slatyer et al. 2013; Saupe et al. 2015), as well as using phylogeographic analyses to explore shifts in geographic range and population size in response to late Quaternary climatic fluctuations (Reid et al. 2006; Rodríguez-Robles et al. 2010; Prates et al. 2016).

Here we compare scales of phylogeographic structure and spatial expansions between a narrowly distributed tropical specialist species of rainbow skink and its widely distributed sister taxon using sequences of mitochondrial DNA (mtDNA) and ~2500 nuclear loci obtained via targeted exon capture (Bragg et al. 2015). Such phylogeographic comparisons of closely related, but ecologically distinct taxa can be highly informative, as other aspects of their biology can be expected to be similar (e.g. Bell et al. 2007; Moussalli et al. 2009). All else being equal, geographically widespread species are more likely to have internal phylogeographic structure as their ranges span multiple biogeographic barriers (Brown et al. 1996; Irwin 2002). However, climatic or habitat specialization will increase sensitivity to past climate change, so that even narrowly distributed species can comprise multiple independently evolving lineages across their range and show more fine scale structure where their ranges overlap with widespread generalists (e.g. Moussalli et al. 2009; Bell et al. 2010). The rainbow skinks, genus *Carlia*, are one of the most speciose groups of skinks in Australia, with most species distributed across northern Australia (Greer et al. 1989; Cogger 2000). This genus has been the subject of comparative studies of speciation and morphological evolution (Phillips et al. 2004; Dolman & Hugall 2008; Dolman & Stuart-Fox 2009; Singhal & Moritz 2012), thermal ecology (Vickers et al. 2011; Goodman et al. 2013; Pintor et al. 2016) and reproductive ecology (James & Shine 1985). Previous phylogeographic analyses of widespread taxa from the monsoonal tropics (Potter et al. 2016) and rainforest taxa from northeast Australia (Dolman & Moritz 2006) have demonstrated strong responses to past

climatic change. However, we still know little about how sympatric and closely related species, with different niche breadths, responded to late Quaternary climatic change.

We focus on *Carlia johnstonei* and *C. triacantha*; sister species that diverged ~ 8 MYA (Dolman & Hugall 2008) and which show contrasting distributions in geographic and climatic space (Fig. 2.1). *C. johnstonei* is restricted to the relatively wet (>750mm annual rainfall) tropical northwest of Australia, where it is the only species of *Carlia* endemic to the Kimberley region. By contrast, *C. triacantha* extends well into the arid zone as well as having a broad distribution across the Australian Monsoonal Tropics (AMT). These differences in geographic range are mirrored in climate space, with *C. johnstonei* occupying a small subset of the climatic range (precipitation and temperature, Fig. 2.1C) of *C. triacantha*. Although little is known about their ecology and behavior, field guides and observations indicate these two leaf-litter species have different habitat preferences, with *C. johnstonei* occurring mostly in tropical woodlands and relatively mesic microenvironments in gullies and bordering creeks (and deeper gorges towards the drier end of their distribution), while *C. triacantha* forages among litter and spinifex mainly on broken stony ground (Cogger 2014; Silva et al., unpublished observations). The closest related species, *C. amax*, occurs across the AMT with geographic and climatic breadth intermediate between the target species, and with multiple phylogeographic lineages (Potter et al. 2016).

The two target species overlap broadly in the remote Kimberley region of northwest Australia, allowing us to test directly how climatic niche breadth may influence demographic history where species likely have experienced the same climatic fluctuations. The Kimberley is a geologically and topographically complex region (Braby 2008; Pepper & Scott Keogh 2014) and experienced strong fluctuations in aridity with changing strength of the Asian Monsoon especially in the early Holocene (Bowman et al. 2010; Reeves et al. 2013; Field et al. 2017). Recent surveys and studies of the previously poorly known fauna have highlighted high endemism in the mesic west Kimberley (Köhler 2010; Morgan et al. 2011; Andersen et al. 2014; Maddock et al. 2015), and deep phylogeographic structure in Kimberley populations of widespread taxa (Oliver et al. 2010; Smith et al. 2011; Potter et al. 2012; Catullo & Scott Keogh 2014; Moritz et al. 2016; Westerman et al. 2017).

Using broad geographic and genomic sampling of the target species, we robustly identify major intraspecific lineages and estimate their divergence histories, examine spatial structuring within lineages and investigate their late Quaternary population dynamics. Our main focus is on comparing species' responses to the most recent (e.g. Holocene) episodes of major climate change. Based on observed differences in geographic and climatic ranges, we expect that the tropical specialist, *C. johnstonei*, will have finer-scale phylogeographic structure than the widespread climate generalist sister species, *C. triacantha*, especially where their ranges overlap in the Kimberley. As well, we expect to infer recent spatial expansion within the specialist, reflecting the stronger demographic fluctuations of a species more sensitive to climate change.

2.2 Materials and Methods

2.2.1 Sampling and laboratory procedures

We obtained most samples from field expeditions in northern and central Australia, and supplemented these with tissues that were available from museum collections (Museum and Art Gallery of the Northern Territory, South Australian Museum, Museum Victoria and Western Australian Museum), to cover the geographic ranges of the studied species (Fig. 1). Using the salting out method of Sunnucks & Hales (1996), we extracted DNA from a total of 162 and 175 samples for *C. triacantha* and *C. johnstonei*, respectively. Two individuals of *C. amax*, the closest taxon to the *C. triacantha* - *C. johnstonei* clade, were used as outgroup samples for phylogenetic analysis and to polarize single nucleotide polymorphisms (SNPs).

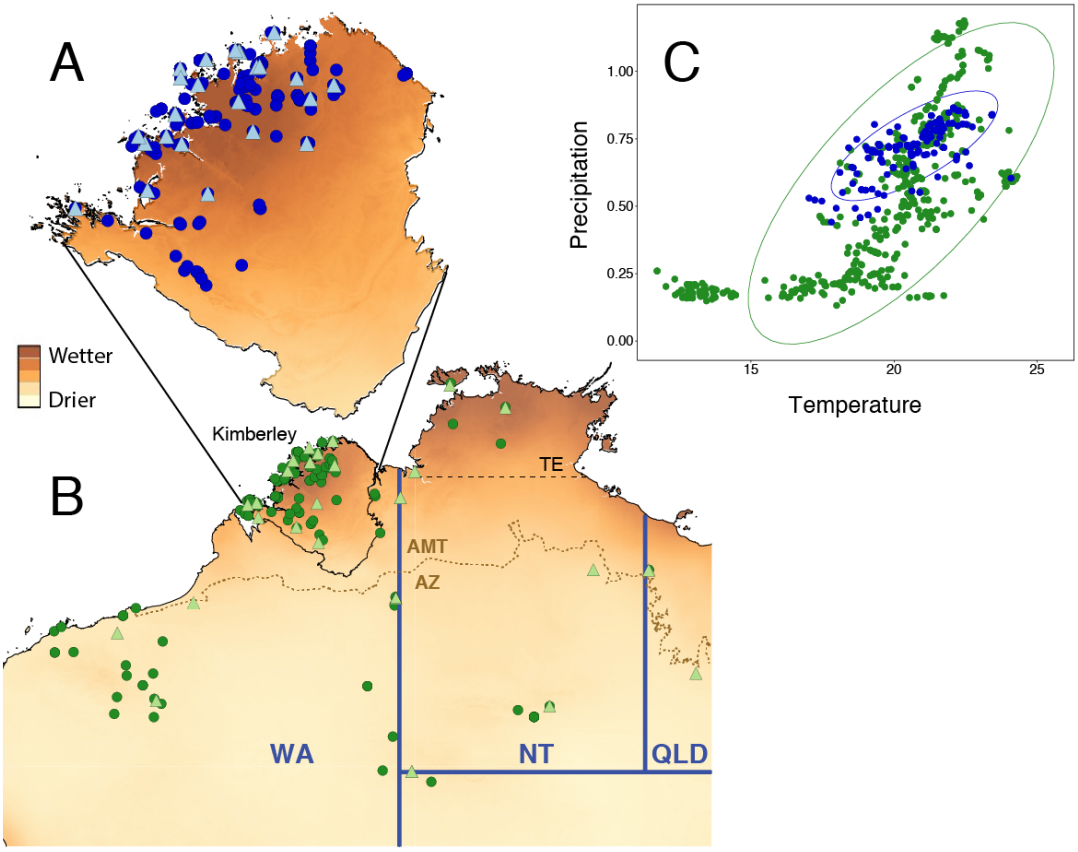


Figure 2.1 Geographic distribution of samples and environmental ranges of (A) the Kimberley restricted tropical specialist *Carlia* johnstonei and (B) of the widespread generalist *Carlia* triacantha on aridity surface (Williams et al. 2010). Circles correspond to samples sequenced for ND4 and triangles to the subset of samples sequenced with the exon capture method. WA, NT, and QLD correspond to Western Australia, Northern Territory and Queensland states; AMT - Australian Monsoonal Tropics and AZ – Arid Zone; TE – Top End. (C) Plot of mean annual precipitation with mean annual temperature (Williams et al. 2010) for both species known locations, showing the difference in climatic niche breadth between *C. johnstonei* (blue dots) and *C. triacantha* (green dots).

We amplified and sequenced the mtDNA gene ND4 using the PCR primers (5'-3'): ND4 light - CACCTATGACTACCAAAAGCTCATGTAGAAGC (Arevalo et al. 1994); Leu3 heavy - GAATTAGCAGTTCTTTRTG (Stuart-Fox et al. 2002), and the protocol described in Potter et

al. (2016). The sequences were edited and aligned in Geneious v.7.1.8 (http://www.geneious.com, Kearse et al. 2012), yielding a final alignment of 722 bp. Sequences are available at NCBI GenBank with accession numbers MF083173 - MF083508.

For exon capture, we selected a subset of 51 samples (Table 2.1) to represent all identified mtDNA lineages, and to span the geographic distribution of each of these lineages. We used a custom, in-solution, exon capture approach (SeqCap EZ Developer Library; Roche NimbleGen). Target probes were designed from multiple transcriptomes of *Carlia* and related genera, as described in Bragg et al. (2015). We prepared genomic libraries following the protocol of Meyer & Kircher (2010) with modifications as in Bi et al. (2012), and we pooled all 51 barcoded libraries in equimolar amounts and hybridized in one sequence capture reaction. The hybridization reaction was executed as in Potter et al. (2016) where, after two independent enrichment PCRs of post-capture libraries (17 cycles), we tested enrichment efficiency by qPCR. We then sequenced the pooled post-capture libraries on a single Illumina HiSeq2000 (100-bp paired-end) lane at the Biomolecular Resource Facility, Australian National University. The raw Illumina reads are archived in the NCBI Short Read Archive, Project ID PRJNA289283, Sample IDs SAMN06927808 - SAMN06927858.

2.2.2 Bioinformatic processing of exon capture data

Illumina sequencing reads were trimmed and cleaned using a workflow developed by Singhal (2013, https://github.com/MVZSEQ). The assembly and mapping of exon targets was performed using a workflow described in Bragg et al. (2015). In brief, we performed an 'exon-specific assembly' for each individual sample. This involved first identifying the sequencing reads with homology to each exon using blastx (Altschul et al. 1990), and then conducting a de novo assembly of those reads. Initial assembly was done using VELVET v.3 (Zerbino & Birney 2008) with a range of K values (K = 31, 41, 51, 61, 71, and 81). The contigs produced by the VELVET assemblies with different K values were then pooled and assembled using CAP3 (Huang & Madan 1999). Exonic regions of the resulting contigs were extracted using EXONERATE v.2.2.0 (Slater et al. 2005). Next, we mapped reads from each sample to the assembled exons using Bowtie2 v.2.2.2 (Langmead & Salzberg 2012). The resulting SAM files were converted to BAM format using SAMtools v.0.1.19 (Li et al. 2009), and read group information was added using Picard v.1.10 (http://picard.sourceforge.net/). We then used GATK v.2.8 (McKenna et al. 2010) to identify heterozygous sites, and we inserted ambiguity codes at appropriate sites in the contig sequence as in Bragg et al. (2015).

The exonic sequences were then aligned and checked for quality using the EAPhy pipeline (Blom 2015; https://github.com/MozesBlom/EAPhy). This pipeline uses MUSCLE v.3.8.31 (Edgar 2004) to align sequences within each locus, performs checks to ensure coding of amino acids, removes missing data from the ends of the sequences, and exports data into FASTA format allowing for missing data (here ranging from 0-10%). Data were output as either a concatenation of all loci (only alignments longer than 150 bp), or just one randomly selected SNP from each locus. We excluded SNPs for which the minor allele was

present in only one copy (singletons), to ensure that SNPs were informative for structure and demographic analyses and to minimize any remaining false SNPs.

Scripts and processed data are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.jj1tt.

2.2.3 Phylogenetic analysis and species tree estimation

We estimated the mtDNA gene tree and the nuclear species tree to explore phylogenetic structure and investigate lineage diversity, before proceeding to analyses of genetic diversity and population structure within each identified lineage.

PartitionFinder v.1.1.1 (Lanfear et al. 2012) was used separately for both ND4 and the concatenated exon capture (allowing 10% individuals as missing data) datasets to select the best-fit partitioning scheme considering partition defined by codon position for each locus. We used the greedy algorithm for mtDNA and the strict hierarchical clustering algorithm for the nuclear dataset, with the model set for RAxML and the Bayesian information criterion (BIC) for model selection in both datasets. The model selected for both datasets was GTR+G+I.

Maximum likelihood (ML) analyses of both datasets were performed using RAxML v.8.0.2 (Stamatakis 2014) with the same settings. Analyses were done using a random starting tree and rapid bootstrap, with clade support estimated via 1,000 bootstrap pseudoreplicates and tree searches from 100 different starting trees.

Taking a representative subset of 20 samples (five per major lineage, and selected so that they represent the diversity within each major lineage) from the exon capture dataset, we used ASTRAL-II v.4.7.6 (Mirarab & Warnow, 2015) to estimate a species tree based on 1643 loci, each with 200bp or longer. This method finds the species tree topology that agrees with the largest number of quartet trees induced by the set of gene trees (Mirarab et al. 2014). Each gene tree was estimated with RAxML, choosing the best tree from 10 runs and conducting 100 bootstrap pseudoreplicates to be used for multi-locus bootstrapping.

2.2.4 Divergence history of major lineages

To infer divergence history among the major lineages within each species, we applied ∂a∂i (v. 1.70, Gutenkunst et al. 2009) which uses a diffusion approximation to fit a likelihood model for demographic history to the two-dimensional-site frequency spectrum (2D-SFS). To achieve a balanced design we used the same samples as for the ASTRAL species tree analyses (i.e. five individuals per lineage), with one randomly selected SNP per locus polarized relative to sequence data from *C. amax*. Specifically, we tested whether an isolation with asymmetric migration model is better supported than a strict isolation model to describe lineage divergences. We performed 50 independent runs with multiple sets of initial parameters, and for model selection we calculated AIC (Akaike Information Criterion) from the run with the best likelihood. The parameters estimated within each species analysis were ancestral effective population size, the effective population size of each daughter lineage,

the time at which the populations split and, for the isolation with migration model, asymmetric migration rates between daughter lineages. To convert parameter estimates from coalescent to absolute units, we assumed a generation time of one year and an exonic mutation rate of 9×10^{-10} as in Singhal & Moritz (2013).

2.2.5 Genetic structure and demographic history within lineages

We assessed spatial structuring of genetic diversity within each of the major lineages, where sample sizes were adequate (n > 10). To maximize the number of loci recovered with high completeness across samples, we rerun the EAPHY separately for each lineage. All analyses to identify spatial clusters were done using one randomly selected SNP from each locus, to avoid bias due to physical linkage, and only SNPs with data for all individuals, so that missing data would not bias results (Patterson et al. 2006).

To identify spatial genetic clusters, we performed an exploratory analysis using a principal coordinate analysis (PCoA) based on Euclidean genetic distances in the package Ade4 (Dray & Dufour 2007) in R version 3.1 (R Core Team 2014). We also used the software STRUCTURE (v.2.3.4, Pritchard et al. 2000) to confirm the same clusters and to estimate the level of admixture within individuals. This analysis was performed using the admixture and independent allele frequencies models, and 10 runs for each value of K between 1 and 5, with a burn-in of 100,000 iterations followed by sampling for 1,000,000 generations. To estimate the most likely K we used the Evanno method (Evanno et al. 2005) in StructureHarvester (v.0.6.94, Earl & VonHoldt 2011). For spatially-explicit analyses, we used the Bayesian assignment program GENELAND (Guillot et al. 2005a; b) and SPACEMIX (v.0.13, Bradburd et al. 2016). Whereas GENELAND uses genetic and geographic information to estimate the most likely number of clusters and their spatial organization, SPACEMIX estimates allele frequency covariances associated with geography and identifies admixture between populations. We conducted five runs in GENELAND to get independent estimates of K (number of populations) and to confirm convergence of the Markov chain Monte Carlo (MCMC) algorithm. Each run was performed using the following parameters: uncorrelated allele frequency model, K between 1 and 5, 10^4 MCMC iterations, thinning interval of 100, and 10 km of uncertainty in the spatial coordinates. In the SPACEMIX analysis, we used the 'source and target' as the long model option, 'counts' as the data type and the analysis was run for 10^7 generations with a thinning of 1×10^4.

Genetic diversity indices and neutrality tests (Tajima's D – Tajima 1989; Fu's Fs – Fu 1997) for mitochondrial lineages and nuclear lineages and spatial clusters were computed in DnaSP v5.10.1 (Librado & Rozas 2009) and in the R package PEGAS (Paradis 2010), respectively. For the nuclear data, we calculated the mean values of the summary and test statistics across loci.

To test for recent (i.e. Holocene) spatial expansion, we used the nuclear dataset and an explicit spatial approach described by Peter & Slatkin (2013) to test lineages with > 10 individuals for population expansion vs. a null model of equilibrium isolation-by-distance.

This method uses skew of the site frequency spectrum to compare equilibrium expected by an isolation-by-distance model against a model in which populations expand on a stepping stone model from a single local population in the recent past. Given evidence of expansion, it measures the strength of the founder effect during the expansion, measured as the effective founder distance: the deme size for which the effective population size during the founder event is reduced by 1%. A low effective founder distance therefore suggests a strong founder effect (Peter & Slatkin 2013). When applying the method, we again used *C. amax* as outgroup to estimate the derived states of the SNPs.

As tests for population size change within lineages can be confounded by past gene flow between lineages (Gattepaille et al. 2013), we also used coalescent simulations in FASTSIMCOAL2 (v.2.5.2.21, Excoffier et al. 2013) and the unfolded 2D-SFS obtained by ∂a∂i, to explore alternative models of divergence and demographic histories (Figure S5). In particular, when allowing for gene flow among lineages, we aimed to contrast models with gradual vs. abrupt changes in population size. As for the ∂a∂i analyses, we compared models with (IM) and without (SI) gene flow, and within each of these, models with a discrete episode of population size change (IM∆N, SI∆N) vs. constant population size (IM, SI). For the IM model, we also compared models with discrete vs. continuous population size change. We focused on model selection rather than parameter estimation as the former was much more robust across independent simulations. For each model, we performed 50 independent runs (10-40 ECM cycles, 1,000,000 simulations per run), chose the run with the highest likelihood for each model, and used AIC for model selection. Given the observed preference for models with discrete change in population size (see results), we then explored models allowing only recent (<10,000 yr) population changes, to reflect expected responses to strong fluctuations in precipitation from the early Holocene (Denniston et al. 2013; Field et al. 2017).

2.3 Results

2.3.1 Mitochondrial DNA phylogeographic patterns

The mtDNA gene tree recovered six well supported lineages (Figs. 2A and S1) with net pairwise divergences ranging from 8% to 18% (Table S1). However, relationships between some of these lineages were unresolved. Individuals morphologically identified as *C. johnstonei* are represented by four strongly supported lineages, forming two divergent clades that are paraphyletic with respect to *C. triacantha*. The lineages Johnstonei A1, A2 and A3, have adjacent distributions in the northwest, northeast and southwest Kimberley, respectively. Within Johnstonei A1, there is an additional north-south phylogeographic break (A1 - N and S) coincident with the Roe River (Figs. 2A and 2C). Johnstonei B is highly divergent from any of the other three conspecific lineages (net divergence >15%; Table S1) and is restricted to the islands of the northwest Kimberley (Fig. 2.2C).

Specimens identified morphologically as *C. triacantha* belong to two widespread and geographically overlapping lineages - Triacantha A and Triacantha B - with net divergence

of ~9% (Fig. 2.2A; Table S2.1). Triacantha A corresponds to the nominal *C. triacantha* and is distributed from western Queensland, across much of the Northern Territory (NT) to the Pilbara in the west (Fig. 2.2D); within this lineage the samples from the mesic northern Top End of the NT are separated both spatially and phylogenetically from those across the arid zone to the south ('mesic' vs. 'arid' lineages; Fig. 2.2A and 2.2D). The second major lineage, Triacantha B, is distributed across the Kimberley and has scattered records eastwards in the NT; no further phylogeographic structure was evident within this lineage.

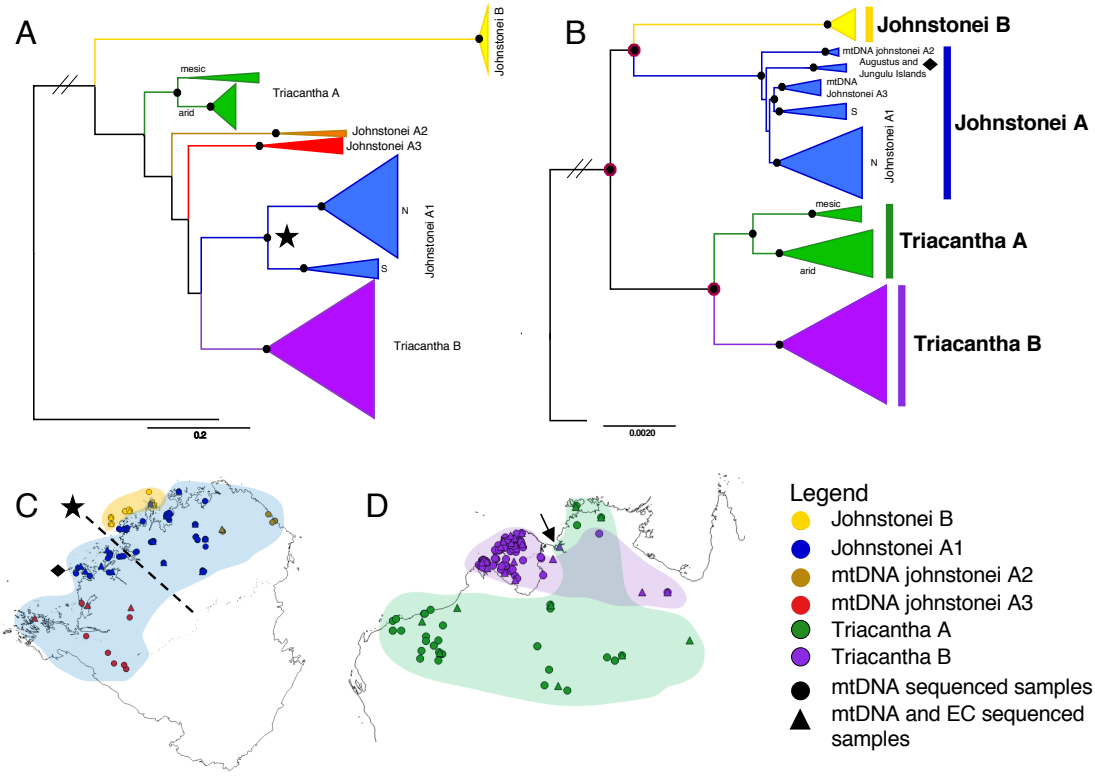


Figure 2.2 Phylogenetic results and geographic distribution of mtDNA and nDNA lineages of *Carlia* johnstonei and *Carlia* triacantha. Top: A) mtDNA ND4 RAxML tree; B) nDNA EC RAxML tree. Black circles represent bootstrap values higher than 70% and black circles with red contour indicate 100% support in the ASTRAL species tree for the major nuclear lineages. Bottom: C) Distribution map of the major mtDNA lineages within *Carlia* johnstonei in the Kimberley; D) Distribution map of the major mtDNA lineages within *C. triacantha*. The shaded areas (blue, yellow, green and purple) correspond to the four major lineages in the nDNA tree. The star and the dashed line (Roe river) signal the north-south mtDNA split within Johnstonei A1, the diamond indicates the Augustus and Jungulu islands and the arrow points to a cytonuclear discordance in *C. triacantha*.

2.3.2 Nuclear phylogeography and divergence histories

The exon capture and sequencing yielded a mean coverage of 40.6x per sample (min=10.70; max=94.89) across an average of 2816 loci per sample (min=1388; max=2961; Table S2.2). The dataset with the two outgroup samples and allowing for 10% missing individuals contained 2423 loci, whereas for no missing data only 480 loci were recovered (mostly due to two lower quality samples, see Table S2.2). All phylogenetic analyses were

done with the dataset allowing for 10% missing individuals, corresponding to a final concatenated alignment of 835,042 bp with 33,597 segregating sites (excluding singletons).

In contrast to the mtDNA phylogeny, the concatenated RAxML phylogenetic tree and the ASTRAL species tree confirmed the reciprocal monophyly of *C. johnstonei* and *C. triacantha* (Fig. 2.2B). Otherwise, the exon data confirmed the phylogenetic separation between the northwestern Kimberley island endemic, Johnstonei B, and the remaining samples (Johnstonei A). Notably, the net nuclear DNA (nDNA) divergence between Johnstonei A and B (0.00123, Table S2.3) is higher than that between Johnstonei A and *C. triacantha* (0.00087; Table S2.3). Within Johnstonei A, the exon data support the phylogenetic distinctiveness of the mtDNA lineages identified above, but also highlight the distinctiveness of samples from central Kimberley islands (Augustus and Jungulu islands; Fig. 2.2C) that for mtDNA were included in the A1(S) lineage. Within *C. triacantha*, the exon-based phylogeny also recovers the Triacantha A and Triacantha B mtDNA clades (Fig. 2.2A), although one geographically intermediate individual has Triacantha B mtDNA with Triacantha A nuclear genome (indicated by arrow in Fig. 2.2D). Further, strong support was found for the mesic-arid split within Triacantha A lineage, but for the nDNA the location of the break was further north than for mtDNA (Fig. S2.2).

Results from the ∂a∂i analyses (Table S2.4) strongly supported a model of isolation-with-migration over strict isolation for the two major lineages of *C. johnstonei* (IM AIC = 2219.63 << SI AIC = 2871.44) and weakly supported the same result for *C. triacantha* (IM AIC = 710.28 < SI AIC = 714.20). However, all estimated migration rates were low (Nm << 1). Estimated average population size for the island endemic Johnstonei B is five-fold smaller than for Johnstonei A, while population sizes for Triacantha A and B are more similar to each other. Regarding divergence times, the split within *C. johnstonei* is older than the split within *C. triacantha*. Absolute divergence times among lineages were estimated at the late Miocene (~5.6 MYA) within *C. johnstonei* and at the late Pliocene (~2.2 MYA) within *C. triacantha*, and are broadly consistent with estimates from mtDNA. Thus, all evidence points to pre-Pleistocene divergences among major intraspecific lineages, substantially predating the late Quaternary population fluctuations inferred from the demographic analyses (see below).

2.3.3 Genetic structure and diversity

To further investigate spatial structure within the major nDNA lineages identified, we applied clustering algorithms to the SNP data for each lineage. Here we present the results for GENELAND and SPACEMIX (Fig. 2.3), while PCoA and STRUCTURE results are shown in supplementary material (Fig. S2.3). Only the Johnstonei A (n=20), Triacantha A (n=10) and Triacantha B (n=16) clades were considered, as sample size for Johnstonei B lineage (n=5) was not adequate for such analyses. These SNP-based analyses support three genetic groups within Johnstonei A: Johnstonei A2 from the northeast Kimberley, samples from the Jungulu and Augustus Islands (off the central Kimberley coast), and the remaining Johnstonei A1 samples. Though GENELAND combined the first two at (the best) K of 2, they were clearly

separated in SPACEMIX (Fig. 2.3) and in the PCoA (Fig. S2.3). In *C. triacantha*, GENELAND and SPACEMIX agreed in supporting two distinct genetic groups within both Triacantha B and Triacantha A (Fig. 2.3), the latter being consistent with the phylogenetic analysis (Fig. 2.2B). No admixture among groups was detected by SPACEMIX in any of the two species.

Clustering using STRUCTURE identified admixture between a northern and a southern cluster within Johnstonei A1 (Fig. S3), with the clusters being similar to the split observed in both mtDNA and nDNA phylogenetic analyses (Fig. 2.2). For Triacantha A and B, STRUCTURE

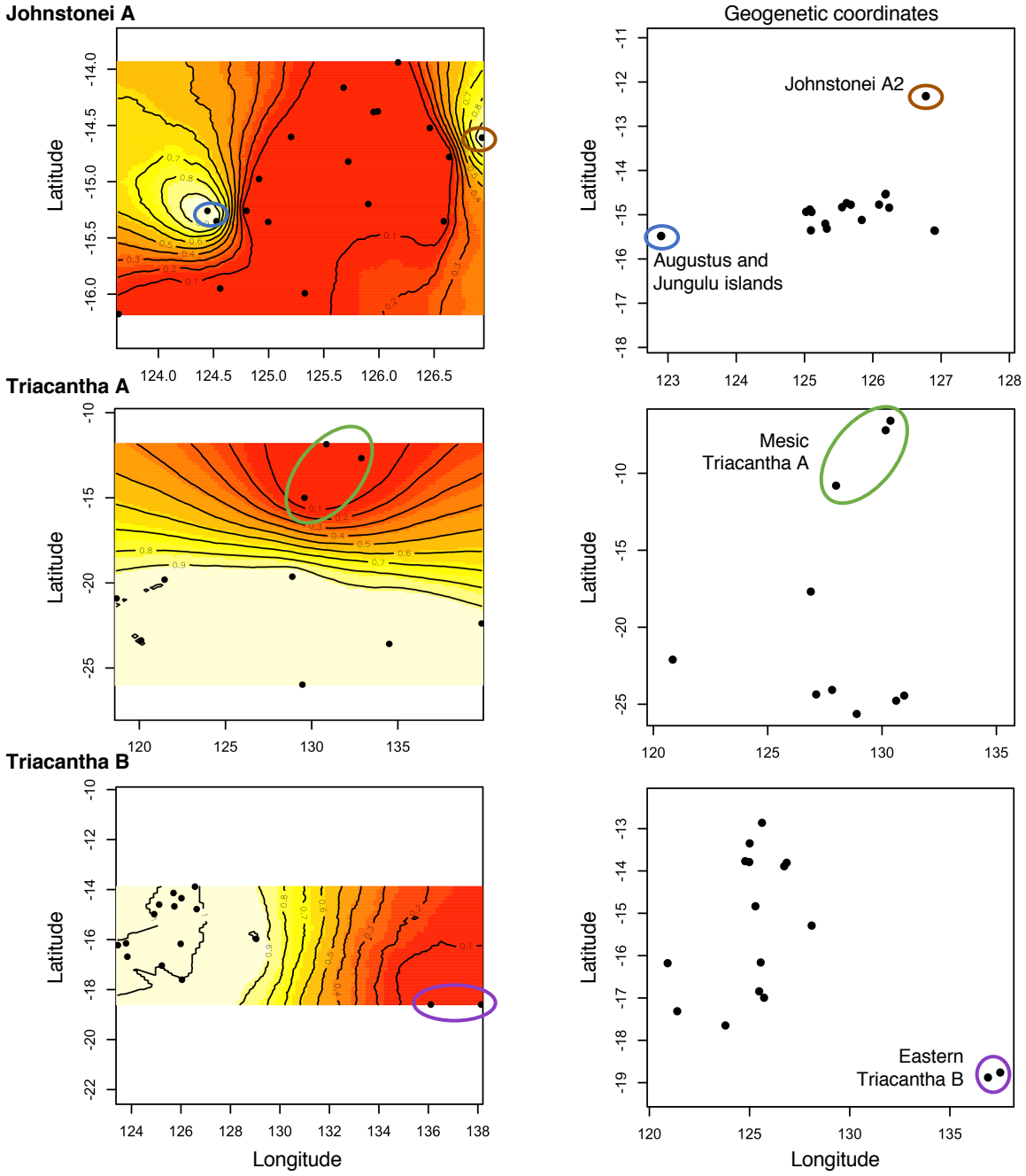


Figure 2.3 Spatial clustering analyses of the nDNA lineages Johnstonei A, Triacantha A and Triacantha B. On the left side are GENELAND heat maps of posterior probability of belonging to cluster one, and on the right side are SPACEMIX geogenetic maps with admixture inference, although no significant admixture was found.

results were consistent with those from the other genetic structure analyses, and did not show evidence of admixture within any of the two lineages. A key finding in both phylogenetic and spatial structure analyses is that, where they overlap in the Kimberley, *C. johnstonei* has strong phylogeographic structure, whereas *C. triacantha* (B) does not. The precipitation with temperature plot (Fig. S2.4) comparing *C. johnstonei* with just the newly discovered Triacantha B lineage shows that the latter lineage also has a broader niche breadth than *C. johnstonei*.

To compare levels of genetic diversity, we focused only on the four major nuclear clades due to sample size limitations. For both mtDNA and nDNA, diversity within Johnstonei B was low (<1%; Tables 2.1, S2.1 and S2.4), which most likely reflects the small insular range of this lineage and lower effective population size inferred from demographic analyses. Johnstonei A had higher overall diversity across exons than the geographically overlapping Triacantha B, as estimated from the number of segregating sites (θ_S) or nucleotide diversity (π) (Table 2.1). In all lineages, θ_S was higher than π, but this difference was greater in Johnstonei A and Triacantha B, and average Tajima's D values were more negative for Johnstonei A and B. Differences in genetic diversity between lineages were apparently not due to the different number of loci considered, since an analysis using the dataset without missing data (480 exons for all samples/lineages) yielded qualitatively similar results (data not shown).

Table 2.1 Genetic diversity indices for each of the main lineages identified in the exon capture phylogenetic tree. Given are sample size (n), number of loci studied, average locus size in base pairs *(bp)*, average theta per site from segregating sites (θ_S), average nucleotide diversity (π), and average Tajima's D per

EC lineages	*n*	Number of loci	*bp*	θ_S	π	Tajima's D
Triacantha A	10	892	479.18	0.00379	0.00172	-2.084
Triacantha B	16	1180	526.47	0.00328	0.00096	-2.158
Johnstonei A	20	2048	442.20	0.00551	0.00165	-2.245
Johnstonei B	5	2496	460.49	0.00081	0.00048	-2.505

2.3.4 Demographic analyses

The Peter & Slatkin (2013) test for spatial range expansion detected a strong and significant deviation from an isolation-by-distance equilibrium in Johnstonei A, but not in either lineage of *C. triacantha* (Table S2.5). The most likely origin of the Johnstonei A expansion is in the north central Kimberley, with an east-west gradient of likelihood possibly reflecting the direction of expansion towards the Kimberley islands (Fig. 2.4). The inferred origin is proximal to the distinct lineage Johnstonei A2. To test whether the divergence of this group within the Johnstonei A clade could have distorted the results, we reran the analysis excluding the Johnstonei A2 samples – there was no substantial change to the results (Table S2.5) or to overall levels of genetic diversity (Table S2.6).

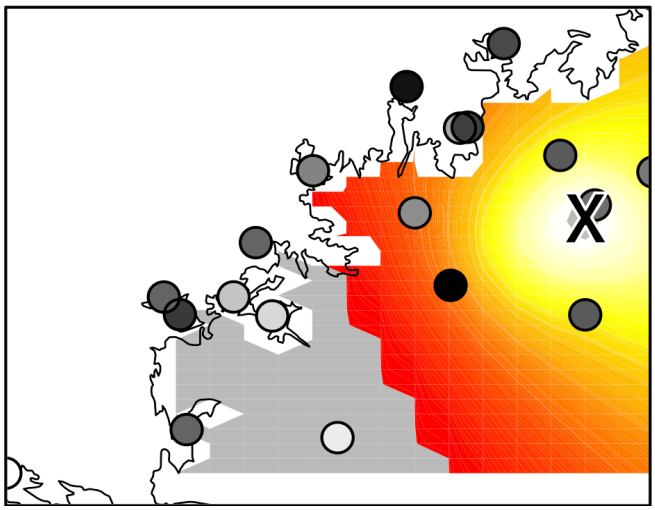


Figure 2.4 Inferred range expansions map for the nDNA lineage Johnstonei A. Samples are represented by filled circles. Light/dark circles indicate high/low heterozygosities. The most likely origin of the expansion is indicated by an X. The yellow shaded area indicates a more likely origin, the red area a less likely origin and the grey area infers regions where the founder effect is negative, and are therefore unlikely to be the source of the expansion.

The FASTSIMCOAL2 analyses strongly and consistently supported models with a combination of migration and population size change for both species (Table 2.2; Figure S2.5). Further, these analyses supported models of discrete over gradual population size changes (with AIC model weights > 0.99). To compare the responses of the two species to the most recent (Holocene) strong climate fluctuations in the Kimberley (i.e. Johnstonei A and Triacantha B lineages), we compared models with discrete population changes allowed in only the last 10,000 years. Though slightly less likely than the unbounded discrete-change models, they were consistent with the results of the spatial expansion tests in supporting recent population expansion for Johnstonei A, whereas Triacantha B appears to have contracted over the same time period (Table 2.2).

Table 2.2 AIC model selection results for FASTSIMCOAL2 analyses. For each species (Sp), the best model is highlighted in italic and the best bounded model for population size change within the last 10,000 years is shown in bold; Par – parameters.

Sp	Model	Maxlog10 (Lhood)	Par	AIC	Weights
C. johnstonei	Strict Isolation (SI)	-4983.35	4	22957.18	0
	Isolation-with-migration (IM)	-4920.19	6	22670.29	1.15E-256
	SI with rapid pop. change (SIΔN)	-4970.76	8	22907.18	4.18E-308
	IM with gradual pop. change	-4822.77	6	22221.68	2.99E-159
	IM with rapid pop. change (IMΔN)	*-4662.51*	*10*	*21491.65*	*1*
	IMΔN with expansion bounded within 10,000 years	**-4668.19**	**10**	**21517.79**	**2.11E-06**
	IMΔN with contraction bounded within 10,000 years	-4669.47	10	21523.71	1.09E-07
C. triacantha	Strict Isolation (SI)	-4319.03	4	19897.85	5.76E-175
	Isolation-with-migration (IM)	-4363.73	6	20107.73	1.53E-220
	SI with rapid pop. change (SIΔN)	-4267.20	8	19667.20	7.02E-125
	IM with gradual pop. change	-4198.12	6	19345.07	6.25E-55
	IM with rapid pop. change (IMΔN)	*-4142.18*	*10*	*19095.45*	*1*
	IMΔN with expansion bounded within 10,000 years	-4146.83	10	19116.86	2.24E-05
	IMΔN with contraction bounded within 10,000 years	**-4143.25**	**10**	**19100.37**	**0.085**

2.4 Discussion

Recent phylogeographic analyses of low dispersal terrestrial vertebrates across the AMT and arid zone have repeatedly revealed deep phylogeographic structure (e.g. Oliver et al. 2010; Potter et al. 2016; Moritz et al. 2016), as we observe here for the widespread *C. triacantha* and its Kimberley-restricted relative, *C. johnstonei*. Before considering the evidence relating to our hypothesis – that the tropical specialist was more sensitive to past climatic change - we first discuss the evidence for multiple lineages within each species in the context of the Kimberley and broader biogeography.

2.4.1 Cryptic lineages

The molecular evidence supports the presence of two deeply divergent lineages in each of *C. johnstonei* and *C. triacantha*. The lineage Johnstonei B is an important new discovery for the region as it is endemic to a set of pristine islands off the northwest Kimberley, which have high conservation value (Gibson & McKenzie 2012). Currently, the Johnstonei B lineage occurs in the Bonaparte Archipelago with confirmed occurrences on the Fenelon, Corneille, East Montalivet, West Montalivet, Don, Berthier, North Maret and South Maret islands, whereas Johnstonei A occurs on other Kimberley islands and the adjacent mainland (Fig. 2.1). As expected given its restricted and insular distribution, diversity and population size for Johnstonei B were much lower than for Johnstonei A. The islands in which Johnstonei B occurs are 'land-bridge islands,' most recently isolated by sea level rise approximately 8,000 to 12,000 years ago (Nix & Kalma 1972), and share much of their species diversity with the mainland Kimberley (Gibson 2014). The high genetic divergence, corresponding to an estimated late Miocene separation, between Johnstonei A and Johnstonei B contrasts with the low divergence among island and mainland populations in other skinks sampled across the same region (within Johnstonei A and Triacantha B, and also Ctenotus inornatus; Harradine et al. 2015). This unexpected pattern of old lineages isolated on young 'land-bridge islands,' has also been observed in the islands off the Top End to the east for the ecologically similar *C. amax* (Potter et al. 2016) and the gecko Heteronotia binoei (Moritz et al. 2016), such that these islands constitute a hotspot of phylogeographic endemism (Rosauer et al. 2016). Comparisons of populations of insects (Francisco et al. 2016) and frogs (Bell et al. 2012) on land-bridge islands with nearby mainland sites in Brazil have also revealed deeper divergences than expected from recent sea level rise, possibly due to intensification of pre-existing genetic differentiation by sea level incursions (Bell et al. 2012). Hawlitschek et al. (2016) also recently found a gecko lineage much older than the estimated geological island age where it occurs.

Within *C. triacantha*, we identified two major lineages that diverged in the late Pliocene or early Pleistocene. The Triacantha B lineage extends from the Kimberley eastwards to the savanna grasslands on the NT/Queensland border, and in the NT overlaps the range of the nominal Triacantha A lineage (mesic Top End and arid zone). By contrast, in other species from the AMT, deep phylogeographic structure typically occurs between the mesic AMT and

the arid zone to the south (Fujita et al. 2010; Melville et al. 2011; Marin et al. 2013) or, within the AMT, between the Kimberley and Top End (Potter et al. 2012, 2016; Huey et al. 2014; Oliver et al. 2014c; Catullo et al. 2014). Unfortunately, samples were not available from the mid-north of the NT (Sturt Plains region) to resolve the distributions of Triacantha A and B where they overlap (Fig. S2.3).

These findings of taxonomically unrecognized diversity in the two *Carlia* species support the recent efforts to better document and understand the rich biodiversity of the Kimberley and the AMT (e.g. Bowman et al. 2010; Doughty 2011; Köhler 2010; Smith et al. 2011; Potter et al. 2012; Palmer et al. 2013; Oliver et al. 2014c), including the National Parks and Indigenous Protected Areas (Moritz et al. 2013) and further emphasize the high endemism of the mesic western Kimberley. Like other recent applications of molecular data to widespread species complexes of vertebrates (e.g. Potter et al. 2014; Rabosky et al. 2014; Oliver et al. 2014a) our observations point to the need to revise the taxonomy of these morphologically similar lineages (Storr 1974), incorporating complementary morphological evidence and appropriate species delimitation analyses.

2.4.2 Climate specialist vs. generalist

The discovery of two major divergent lineages within *C. triacantha* allows for a more geographically matched comparison, now between the largely Kimberley-restricted Triacantha B lineage and the Johnstonei A lineage. As for *C. triacantha* as a whole, the Triacantha B lineage occupies a much broader climatic space across its Kimberley distribution than does *C. johnstonei* (Fig. S2.4). Consistent with our initial expectations, the Johnstonei A lineage showed clear phylogeographic structure and evidence of range expansion, the latter supported by both the spatially explicit test for range expansion and the FASTSIMCOAL2 analyses. Within Johnstonei A, strong subdivision was identified between the northeast (lineage Johnstonei A2) and northwest Kimberley - corresponding to distinct sandstone plateaus separated by low-lying basalt plains, and also between the mainland and the relatively large Augustus island and nearby Jungulu island (Figs. 2C and 3). By comparison, for Triacantha B in the Kimberley, there was no obvious internal phylogeographic structure and there was no significant deviation from isolation-by-distance using the spatially explicit test of Peter and Slatkin (2013). Though the FASTSIMCOAL2 analyses supported models with a discrete population size change sometime in the past for both species, models with size change restricted to the recent past indicated opposite responses, with *C. johnstonei* expanding and *C. triacantha* contracting. Though we are unable to resolve the exact timing of these events, it is possible that they were associated with the early-mid Holocene increase in monsoonal precipitation which would have favoured the more mesic *C. johnstonei*.

Given that the geographically overlapping Johnstonei A and Triacantha B lineages have experienced a common history of late Quaternary climatic change, our findings are consistent with the more specialized Johnstonei A having responded more strongly to past climatic change. In support, a parallel analysis of *C. amax*, which has climatic niche breadth

and range size intermediate between *C. johnstonei* and *C. triacantha*, revealed range expansion from the Kimberley towards the western Top End, but no strong phylogeographic structure within the Kimberley (Potter et al. 2016). An alternative explanation is that the stronger geographical structuring across the range of *C. johnstonei* is simply due to restricted dispersal across the geologically complex Kimberley landscape in this mesic habitat-associated species (e.g. Pepper et al. 2011; Oliver et al. 2014b). This scenario is plausible for land snails of the Kimberley which show extensive short-range endemism of species (Köhler 2010). However, for *C. johnstonei*, limited dispersal alone is not consistent with the inferred history of spatial expansion of the Johnstonei A lineage.

Tropical ectotherm species are known for high sensitivity to climate change, but not all have the same susceptibility, varying in climatic range and habitat restriction. Most studies comparing habitat or resource specialist vs. generalist species have shown stronger genetic structure in the specialist species (e.g. Kelley et al. 2000; Brouat et al. 2003; Eberhart-Phillips et al. 2015; Dellicour et al. 2015). This suggests that specialist species either have generally lower dispersal rates and/or smaller meta-population sizes than generalists, or that they have experienced stronger population reductions (Kelley et al. 2000), likely due to past environmental change. Specialists can be potentially more vulnerable to climatic changes largely due to intrinsic physiological sensitivity, lower dispersal capacity, or both.

A similar comparative study in *Lampropholis* skinks (sister genus to *Carlia*) from the Australian Wet Tropics showed that a montane specialist is more strongly structured than a more broadly distributed congener and that populations have persisted in multiple refugia through past climatic change (Bell et al. 2010), possibly because of strong plasticity of both physiological limits (Llewelyn et al. 2016) and thermoregulatory behaviour (as in other *Carlia* species; Vickers et al. 2011). Experimental studies in Drosophila found that narrowly distributed tropical species consistently have lower genetic variance for climatic related traits as compared with those of widely distributed species (Kellermann et al. 2009). These observations, together with our results, suggest that specialist species are likely to be constrained in their evolutionary responses to future climate changes. For such taxa, we now need to move beyond descriptions of climate space and phylogeography to develop physiology-based mechanistic models of responses to past and future climate change (Kearney & Porter 2009).

2.5 References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–10.

Andersen AN, Bocciarelli D, Fairman R, Radford IJ (2014) Conservation status of ants in an iconic region of monsoonal Australia: levels of endemism and responses to fire in the eastern Kimberley. *Journal of Insect Conservation*, **18**, 137–146.

Arevalo E, Davis SK, Sites JW (1994) Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Biology*, **43**, 387–418.

Bell RC, Brasileiro CA, Haddad CFB, Zamudio KR (2012) Evolutionary history of *Scinax* treefrogs on land-bridge islands in south-eastern Brazil. *Journal of Biogeography*, **39**, 1733–1742.

Bell KL, Moritz C, Moussalli A, Yeates DK (2007) Comparative phylogeography and speciation of dung beetles from the Australian Wet Tropics rainforest. *Molecular Ecology*, **16**, 4984–4998.

Bell RC, Parra JL, Tonione M et al. (2010) Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Molecular Ecology*, **19**, 2531–44.

Berkman LK, Nielsen CK, Roy CL, Heist EJ (2015) Comparative genetic structure of sympatric leporids in southern Illinois. *Journal of Mammalogy*, **96**, 552–563.

Bi K, Vanderpool D, Singhal S et al. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC genomics*, **13**, 403.

Blom MPK (2015) EAPhy: A flexible tool for high-throughput quality filtering of exon-alignments and data processing for phylogenetic methods. *PLoS Currents Tree of Life*, 1.

Bowman DMJS, Brown GK, Braby MF et al. (2010) Biogeography of the Australian monsoon tropics. *Journal of Biogeography*, **37**, 201–216.

Braby MF (2008) Biogeography of butterflies in the Australian monsoon tropics. *Australian Journal of Zoology*, **56**, 41.

Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. *PLoS Genetics*, **12**, e1005703.

Bragg JG, Potter S, Bi K, Moritz C (2015) Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources*, **16**, 1059–1068.

Brouat C, Sennedot F, Audiot P, Leblois R, Rasplus JY (2003) Fine-scale genetic structure of two carabid species with contrasted levels of habitat specialization. *Molecular Ecology*, **12**, 1731–1745.

Brown JH, Stevens GC, Kaufman DM (1996) The Geographic Range: Size, Shape, Boundaries, and Internal Structure. *Annual Review of Ecology and Systematics*, **27**, 597–623.

Catullo RA, Lanfear R, Doughty P, Keogh JS (2014) The biogeographical boundaries of northern Australia: evidence from ecological niche models and a multi-locus phylogeny of *Uperoleia* toadlets (Anura: Myobatrachidae). *Journal of Biogeography*, **41**, 659–672.

Catullo R, Scott Keogh J (2014) Aridification drove repeated episodes of diversification between Australian biomes: Evidence from a multi-locus phylogeny of Australian toadlets (*Uperoleia*: Myobatrachidae). *Molecular Phylogenetics and Evolution*, **79**, 106–117.

Cogger H (2014) *Reptiles and amphibians of Australia*. New Holland, Sydney.

Dellicour S, Michez D, Mardulyn P (2015) Comparative phylogeography of five bumblebees: impact of range fragmentation, range size and diet specialization. *Biological Journal of the Linnean Society*, **116**, 926–939.

Denniston RF, Asmerom Y, Lachniet M et al. (2013) A Last Glacial Maximum through middle Holocene stalagmite record of coastal Western Australia climate. *Quaternary Science Reviews*, **77**, 101–112.

Deutsch CA, Tewksbury JJ, Huey RB et al. (2008) Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 6668–72.

Dolman G, Hugall AF (2008) Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Molecular Phylogenetics and Evolution*, **49**, 782–94.

Dolman G, Moritz C (2006) A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution*, **60**, 573–582.

Dolman G, Stuart-Fox D (2009) Processes driving male breeding colour and ecomorphological diversification in rainbow skinks: a phylogenetic comparative test. *Evolutionary Ecology*, **24**, 97–113.

Doughty P (2011) An emerging frog diversity hotspot in the northwest Kimberley of Western Australia: another new frog species from the high rainfall zone. *Records of the Western Australian Museum*, **26**, 209–216.

Dray S, Dufour AB (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, **22**, 1–20.

Earl DA, VonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eberhart-Phillips LJ, Hoffman JI, Brede EG et al. (2015) Contrasting genetic diversity and population structure among three sympatric Madagascan shorebirds: parallels with rarity, endemism, and dispersal. *Ecology and Evolution*, **5**, 997–1010.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–7.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–20.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, **9**, e1003905.

Field E, McGowan HA, Moss PT, Marx SK (2017) A late Quaternary record of monsoon variability in the northwest Kimberley, Australia. *Quaternary International*, **449**, 119-135.

Francisco FO, Santiago LR, Mizusawa YM, Oldroyd BP, Arias MC (2016) Genetic structure of island and mainland populations of a Neotropical bumble bee species. *Journal of Insect Conservatio*n, **20**, 383–394.

Fu YX (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics*, **147**, 915–925.

Fujita MK, McGuire JA, Donnellan SC, Moritz C (2010) Diversification and persistence at the arid-monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (*Heteronotia binoei*; Gekkonidae). *Evolution*, **64**, 2293–314.

Gattepaille LM, Jakobsson M, Blum MG (2013) Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity*, **110**, 409–419.

Gibson LA (2014) Biogeographic patterns on Kimberley islands, Western Australia. *Records of the Western Australian Museum Supplement*, **81**, 245–280.

Gibson L, McKenzie NL (2012) Identification of biodiversity assets on selected Kimberley islands: background and implementation. *Records of the Western Australian Museum Supplement*, **81**, 1–14.

Goodman B a, Schwarzkopf L, Krockenberger AK (2013) Phenotypic integration in response to incubation environment adaptively influences habitat choice in a tropical lizard. *The American naturalist*, **182**, 666–73.

Greer AE (1989) *The biology and evolution of Australian lizards*. Surrey Beatty and Sons, Sydney.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–80.

Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genetics*, **5**, e1000695.

Harradine EL, Andrew ME, Thomas JW et al. (2015) Importance of dispersal routes that minimize open-ocean movement to the genetic structure of island populations. *Conservation Biology*, **29**, 1704–1714.

Hawlitschek O, Toussaint EFA, Gehring P-S et al. (2017) Gecko phylogeography in the Western Indian Ocean region: the oldest clade of Ebenavia inunguis lives on the youngest island. *Journal of Biogeography*, **44**, 409–420.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome research*, **9**, 868–77.

Huey JA, Cook BD, Unmack PJ, Hughes JM (2014) Broadscale phylogeographic structure of five freshwater fishes across the Australian Monsoonal Tropics. *Freshwater Science*, **33**, 273–287.

Huey RB, Deutsch CA, Tewksbury JJ et al. (2009) Why tropical forest lizards are vulnerable to climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 1939–48.

Huey RB, Kearney MR, Krockenberger A et al. (2012) Predicting organismal vulnerability to climate warming: roles of behaviour, physiology and adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 1665–79.

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, **56**, 2383–94.

James C, Shine R (1985) The seasonal timing of reproduction: A tropical-temperate comparison in Australian lizards. *Oecologia*, **67**, 464–474.

Janzen DH (1967) Why mountain passes are higher in the tropics. *The American Naturalist*, **101**, 233–249.

Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, **25**, 185–202.

Kearney MR, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology letters*, **12**, 334–50.

Kearse M, Moir R, Wilson A et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–9.

Kellermann V, van Heerwaarden B, Sgrò CM, Hoffmann AA (2009) Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science*, **325**, 1244–6.

Kelley ST, Farrell BD, Mitton JB (2000) Effects of specialization on genetic differentiation in sister species of bark beetles. *Heredity*, **84**, 218.

Köhler F (2010) Uncovering local endemism in the Kimberley, Western Australia: description of new species of the genus *Amplirhagada* Iredale, 1933 (Pulmonata: Camaenidae). *Records of the Australian Museum*, **62**, 217–284.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–701.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–9.

Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–9.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Llewelyn J, Macdonald SL, Hatcher A, Moritz C, Phillips BL (2016) Intraspecific variation in climate-relevant traits in a tropical rainforest lizard. *Diversity and Distributions*, **22**, 1000–1012.

López-Uribe MM, Zamudio KR, Cardoso CF, Danforth BN (2014) Climate, physiological tolerance and sex-biased dispersal shape genetic structure of Neotropical orchid bees. *Molecular Ecology*, **23**, 1874–90.

Maddock ST, Ellis RJ, Doughty P, Smith LA, Wüster W (2015) A new species of death adder (Acanthophis: Serpentes: Elapidae) from north-western Australia. *Zootaxa*, **4007**, 301–26.

Marin J, Donnellan SC, Hedges SB et al. (2013) Hidden species diversity of Australian burrowing snakes (*Ramphotyphlops*). *Biological Journal of the Linnean Society*, **110**, 427–441.

McCain CM (2009) Vertebrate range sizes indicate that mountains may be "higher" in the tropics. *Ecology letters*, **12**, 550–60.

McDonough MM, Šumbera R, Mazoch V et al. (2015) Multilocus phylogeography of a widespread savanna-woodland-adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region. *Molecular Ecology*, **24**, 5248–5266.

McKenna A, Hanna M, Banks E et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**, 1297–303.

Melville J, Ritchie EG, Chapple SNJ, Glor RE, Schulte JA (2011) Evolutionary origins and diversification of dragon lizards in Australia's tropical savannas. *Molecular Phylogenetics and Evolution*, **58**, 257–270.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **5**.

Mirarab S, Reaz R, Bayzid MS et al. (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**, i541-8.

Mirarab S, Warnow T (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, **31**, i44–i52.

Morgan DL, Allen GR, Pusey BJ, Burrows DW (2011) A review of the freshwater fishes of the Kimberley region of Western Australia. *Zootaxa*, **1**, 1–64.

Moritz C, Ens EJ, Potter S, Catullo RA (2013) The Australian monsoonal tropics: An opportunity to protect unique biodiversity and secure benefits for Aboriginal communities. *Pacific Conservation*, **19**, 343–355.

Moritz C, Fujita MK, Rosauer D et al. (2016) Multilocus phylogeography reveals nested endemism in a gecko across the monsoonal tropics of Australia. *Molecular Ecology*, **25**, 1354–1366.

Moussalli A, Moritz C, Williams SE, Carnaval AC (2009) Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. *Molecular Ecology*, **18**, 483–99.

Nix HA, Kalma JD (1972) *Bridge and Barrier: The Natural and Cultural History of Torres Strait*. Australian National University Press, Canberra.

Oliver PM, Adams M, Doughty P (2010) Molecular evidence for ten species and Oligo-Miocene vicariance within a nominal Australian gecko species (*Crenadactylus ocellatus*, Diplodactylidae). *BMC Evolutionary Biology*, **10**, 386.

Oliver P, Couper PJ, Pepper M (2014a) Independent Transitions between Monsoonal and Arid Biomes Revealed by Systematic Revison of a Complex of Australian Geckos (*Diplodactylus*; Diplodactylidae). *PLoS ONE*, **9**, e111895.

Oliver PM, Laver RJ, Melville J, Doughty P (2014b) A new species of Velvet Gecko (*Oedura*: Diplodactylidae) from the limestone ranges of the southern Kimberley, Western Australia. *Zootaxa*, **3873**, 49–61.

Oliver PM, Laver RJ, Smith KL, Bauer AM (2014c) Long-term persistence and vicariance within the Australian Monsoonal Tropics: the case of the giant cave and tree geckos (*Pseudothecadactylus*). *Australian Journal of Zoology*, **61**, 462–468.

Palmer R, Pearson DJ, Cowan MA, Doughty P (2013) Islands and scales: a biogeographic survey of reptiles on Kimberley islands, Western Australia. *Records of the Western Australian Museum*, **204**, 183–204.

Paradis E (2010) Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419–20.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS genetics*, **2**, e190.

Pepper M, Fujita MK, Moritz C, Keogh JS (2011) Palaeoclimate change drove diversification among isolated mountain refugia in the Australian arid zone. *Molecular Ecology*, **20**, 1529–45.

Pepper M, Scott Keogh J (2014) Biogeography of the Kimberley, Western Australia: a review of landscape evolution and biotic response in an ancient refugium. *Journal of Biogeography*, **41**, 1443–1455.

Peres EA, Sobral-Souza T, Perez MF et al. (2015) Pleistocene niche stability and lineage diversification in the subtropical spider *Araneus omnicolor* (Araneidae). *PloS one*, **10**, e0121543.

Peter BBM, Slatkin M (2013) Detecting range expansions from genetic data. *Evolution*, **67**, 3274–3289.

Phillips BL, Baird SJE, Moritz C (2004) When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, **58**, 1536–1548.

Pincheira-Donoso D (2011) Predictable variation of range-sizes across an extreme environmental gradient in a lizard adaptive radiation: evolutionary and ecological inferences. *PloS one*, **6**, e28942.

Pintor AF V, Schwarzkopf L, Krockenberger AK (2016) Extensive acclimation in ectotherms conceals interspecific variation in thermal tolerance limits. *PloS one*, **11**, e0150408.

Pither J (2003) Climate tolerance and interspecific variation in geographic range size. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 475–81.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, **25**, 1367–1380.

Potter S, Close RL, Taggart DA, Cooper SJB, Eldridge MDB (2014) Taxonomy of rock-wallabies, Petrogale (Marsupialia: Macropodidae). IV. Multifaceted study of the *brachyotis* group identifies additional taxa. Australian Journal of Zoology, **62**, 401–414.

Potter S, Eldridge MDB, Taggart DA, Cooper SJB (2012) Multiple biogeographical barriers identified across the monsoon tropics of northern Australia: phylogeographic analysis of the *brachyotis* group of rock-wallabies. *Molecular Ecology*, **21**, 1–16.

Prates I, Xue AT, Brown JL et al. (2016) Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences*, **113**, 7978–7985.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945–959.

Rabosky DL, Hutchinson MN, Donnellan SC, Talaba AL, Lovette IJ (2014) Phylogenetic disassembly of species boundaries in a widespread group of Australian skinks (Scincidae: *Ctenotus*). *Molecular Phylogenetics and Evolution*, **77**, 71–82.

Reeves JM, Bostock HC, Ayliffe LK et al. (2013) Palaeoenvironmental change in tropical Australasia over the last 30,000 years – a synthesis by the OZ-INTIMATE group. *Quaternary Science Reviews*, **74**, 97–114.

Reid DG, Lal K, Mackenzie-Dodds J et al. (2006) Comparative phylogeography and species boundaries in Echinolittorina snails in the central Indo-West Pacific. *Journal of Biogeography*, **33**, 990–1006.

Rödder D, Lawing AM, Flecks M et al. (2013) Evaluating the significance of paleophylogeographic species distribution models in reconstructing Quaternary range-shifts of Nearctic chelonians. *PLoS ONE*, **8**, e72855.

Rodríguez-Robles JA, Jezkova T, Leal M (2010) Climatic stability and genetic divergence in the tropical insular lizard *Anolis krugi*, the Puerto Rican "Lagartijo Jardinero de la Montaña." *Molecular Ecology*, **19**, 1860–76.

Rosauer DF, Blom MPK, Bourke G et al. (2016) Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia. *Biological Conservation*, **204**, 83–93.

Saupe EE, Qiao H, Hendricks JR et al. (2015) Niche breadth and geographic range size as determinants of species survival on geological time scales. *Global Ecology and Biogeography*, **24**, 1159–1169.

Singhal S (2013) De novo transcriptomic analyses for non-model organisms: An evaluation of methods across a multi-species data set. *Molecular Ecology Resources*, **13**, 403–416.

Singhal S, Moritz C (2012) Testing hypotheses for genealogical discordance in a rainforest lizard. *Molecular Ecology*, **21**, 5059–72.

Singhal S, Moritz C (2013) Reproductive isolation between phylogeographic lineages scales with divergence. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20132246.

Slater GSC, Birney E, Box G et al. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.

Slatyer RA, Hirst M, Sexton JP (2013) Niche breadth predicts geographical range size: a general ecological pattern. *Ecology Letters*, **16**, 1104–14.

Smith KL, Harmon LJ, Shoo LP, Melville J (2011) Evidence of constrained phenotypic evolution in a cryptic species complex of agamid lizards. *Evolution*, **65**, 976–992.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–3.

Storr GM (1974) The genus *Carlia* (Lacertilia, Scincidae) in Western Australia and Northern Territory. *Records of the Western Australian Museum*, **3**, 151–165.

Stuart-Fox DM, Hugall AF, Moritz C (2002) A molecular phylogeny of rainbow skinks (Scincidae: *Carlia*): taxonomic and biogeographic implications. *Australian Journal of Zoology*, **50**, 39–51.

Sunday JM, Bates AE, Dulvy NK et al. (2011) Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1823–30.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510–524.

Tajima F (1989) Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, **123**, 585–595.

Vickers M, Manicom C, Schwarzkopf L (2011) Extending the cost-benefit model of thermoregulation: high-temperature environments. *The American Naturalist*, **177**, 452–61.

Werneck FP, Costa GC, Colli GR, Prado DE, Sites Jr JW (2011) Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. *Global Ecology and Biogeography*, **20**, 272–288.

Westerman M, Blacket MJ, Hintz A et al. (2017) A plethora of *planigales*: genetic variability and cryptic species in a genus of dasyurid marsupials from northern Australia. *Australian Journal of Zoology*, **64**, 303–311.

Williams K, Ferrier S, Rosauer D et al. (2010) *Harnessing continent-wide biodiversity datasets for prioritising national conservation investment*. A report to the Australian Government Department of the Environment, Water, Heritage and the Arts.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, **18**, 821–9.

Acknowledgements

We thank the Museum and Art Gallery of the Northern Territory, Museum Victoria, South Australia Museum, Western Australia Museum, the Australian Wildlife Conservancy and Dunkeld Pastoral Company for samples. And to all the people that helped with fieldwork to expand the sampling available for this study.

We are also grateful to anonymous reviewers for helpful comments, Sonal Singhal and Alexander Xue assistance with demographic analysis and to members of the Moritz lab for ongoing discussions relating to this study, and in particular to Mozes Blom, Joshua Peñalba, Huw Ogilvie, Rebecca Laver, Dan Rosauer and Maria Strangas for advice on analysis and comments on the manuscript. A.C.A.S is supported by the FCT grant SFRH/BD/88740/2012 and this research was supported by an ARC Laureate Fellowship awarded to C.M. (ARC FL110100104). This work used the ACRF Biomolecular Resource Facility of the Australian National University.

Data Accessibility

Mitochondrial ND4 sequences associated with this study are available at NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/; Accession numbers MF083173 - MF083508). Raw sequencing reads associated with this study are available at NCBI Sequence Read Archive (http:// www.ncbi.nlm.nih.gov/sra; Project Accession: BioProject PRJNA289283; BioSamples SAMN06927808 - SAMN06927858). ND4 GenBank samples list, exon capture code and datasets that were used in data analysis are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.jj1tt.

Author Contributions

A.C.A.S. designed the research, did the laboratorial work, data analyses, and wrote the manuscript. J.G.B. did the initial bioinformatics, wrote a methods section of the manuscript and commented on the rest. S.P. assisted with laboratorial work, advised in data analyses and commented on the manuscript. M.M.C. and C.F. had intellectual input into the direction of this study and commented on the manuscript. C.M. designed the research, assisted in writing the manuscript and funded the work.

2.6 Supplementary material

Table S2.1 Average number of nucleotide substitutions (Dxy; above diagonal) and number of net nucleotide substitutions (Da) per site (below diagonal) between identified mtDNA lineages. Values in bold along the diagonal are the average number of nucleotide differences within lineages. Calculated with DnaSP (v5.10.1, Librado & Rozas 2009).

	Triacantha A	Triacantha B	Johnstonei B	Johnstonei A1	Johnstonei A2	Johnstonei A3
Triacantha A	**0.03979**	0.14926	0.17727	0.13747	0.13235	0.12059
Triacantha B	0.09165	**0.08845**	0.17986	0.14703	0.14809	0.13401
Johnstonei B	0.15497	0.13775	**0.00686**	0.18752	0.19028	0.18453
Johnstonei A1	0.08537	0.07511	0.15091	**0.07236**	0.15386	0.13932
Johnstonei A2	0.11047	0.10639	0.18389	0.11766	**0.04664**	0.13378
Johnstonei A3	0.08655	0.08015	0.16599	0.09096	0.11565	**0.03156**

Table S2.2 Exon capture sample list with tissue number, location, exon capture (EC) lineages, number of reads, mean coverage and number of loci for each sample. For tissue numbers, R refers to the West Australian Museum herpetology collection, ABTC to the Australian Biological Tissue Collection at the South Australian Museum, and CCM to the Moritz lab collection at The Australian National University.
*Samples used in ASTRAL and FASTSIMCOAL2 analyses.

Individual	Tissue number	Latitude	Longitude	EC lineage	Total number of reads	Mean coverage for each sample	Number of loci
AS01_indexing1	R117946	-14.137778	125.698611	Triacantha_B	972321	14.11	2323
AS01_indexing2	R158008	-16.148611	123.779444	Triacantha_B	2006372	42.25	2933
AS01_indexing3	R168462	-13.883333	126.566667	Triacantha_B	2312358	25.91	2772
AS01_indexing4	R171912	-16.2225	123.449722	Triacantha_B	751329	10.70	1556
AS01_indexing5*	ABTC29851	-15.970833	129.040278	Triacantha_B	3325552	55.59	2946
AS01_indexing6*	ABTC61613	-15.000000	129.583300	Triacantha_A	1758912	29.62	2905
AS01_indexing7*	CCM1130	-14.781100	126.634900	Triacantha_B	2864826	53.87	2951
AS01_indexing8	CCM1202	-14.167100	125.733200	Triacantha_B	2027848	35.20	2934
AS01_indexing9	CCM1235	-17.040670	125.226800	Triacantha_B	2412697	47.20	2933
AS01_indexing10	CCM1643	-16.171080	125.987630	Triacantha_B	2092954	53.58	2932
AS01_indexing11	CCM1767	-17.605600	126.041100	Triacantha_B	1256323	23.21	2881
AS01_indexing12*	ABTC28797	-18.600000	136.100000	Triacantha_B	2536031	45.51	2934
AS01_indexing13*	ABTC29531	-18.600000	138.130000	Triacantha_B	2487031	46.80	2934
AS01_indexing14	R117798	-14.343300	126.020800	Triacantha_B	1277710	26.81	2893
AS01_indexing15	R146022	-16.683300	123.833300	Triacantha_B	1958205	31.18	2900
AS01_indexing16*	R168590	-14.983333	124.916667	Triacantha_B	3769802	56.94	2941
AS01_indexing17	R168910	-14.600000	125.116700	Triacantha_B	2538412	40.28	2951
AS01_indexing18	R117787	-14.380278	125.951111	Johnstonei_A	1953522	30.35	2929
AS01_indexing19	R158783	-14.601944	125.203889	Johnstonei_A	2439110	40.82	2946
AS01_indexing20*	BP00664	-14.976300	124.913100	Johnstonei_A	2920719	49.35	2947
AS01_indexing21	R171223	-15.259167	124.799722	Johnstonei_A	1311975	18.19	2717
AS01_indexing22	R171227	-15.351389	124.526944	Johnstonei_A	1928136	31.54	2932
AS01_indexing23	R171243	-15.358056	124.996111	Johnstonei_A	1503743	20.92	2843
AS01_indexing24	R171897	-13.939722	126.173611	Johnstonei_A	1612585	18.70	2567
AS01_indexing25	CCM0737	-15.198900	125.903500	Johnstonei_A	2580074	50.38	2946
AS01_indexing26	CCM0762	-15.352200	126.588100	Johnstonei_A	834857	19.86	2816
AS01_indexing27*	CCM1125	-14.780600	126.636900	Johnstonei_A	2312068	62.49	2937
AS01_indexing28	CCM1196	-14.821100	125.721400	Johnstonei_A	1789957	47.76	2915
AS01_indexing29*	R171237	-15.259722	124.445278	Johnstonei_A	2929753	62.85	2961
AS01_indexing30	CCM0934	-14.522050	126.461210	Johnstonei_A	4195936	92.25	2937
AS01_indexing31	R117726	-14.163800	125.679700	Johnstonei_A	3722703	91.23	2944
AS01_indexing32	R117838	-14.375500	125.989700	Johnstonei_A	1429716	33.11	2922
AS01_indexing33	R110905	-19.632222	128.875278	Triacantha_A	839914	12.48	1388
AS01_indexing34*	R139010	-19.808333	121.463889	Triacantha_A	2833518	44.78	2937
AS01_indexing35	R154566	-23.379167	120.105278	Triacantha_A	1377935	17.48	2083
AS01_indexing36	ABTC113761	-22.385000	139.856389	Triacantha_A	1407970	21.22	2858
AS01_indexing37	ABTC30066	-11.852222	130.852500	Triacantha_A	1883701	27.83	2920
AS01_indexing38*	ABTC41790	-25.980278	129.463056	Triacantha_A	1766697	25.96	2885
AS01_indexing39*	ABTC29091	-12.670000	132.880000	Triacantha_A	3436297	49.80	2932
AS01_indexing40	R145556	-20.910000	118.680000	Triacantha_A	1003340	13.88	2262
AS01_indexing41*	CCM1859	-23.583380	134.499420	Triacantha_A	2901820	40.58	2930
AS01_indexing42*	R158583	-14.279444	125.306111	Johnstonei_B	2887971	34.91	2920
AS01_indexing43*	R158657	-14.516667	124.983333	Johnstonei_B	1520389	19.62	2803
AS01_indexing44*	R158705	-14.398333	124.977500	Johnstonei_B	1999583	27.26	2913
AS01_indexing45*	R117953	-14.185800	125.734700	Johnstonei_B	3881197	87.72	2933
AS01_indexing46*	R117967	-14.185500	125.724400	Johnstonei_B	5322885	89.91	2947
AS01_indexing47*	R173481	-14.608592	126.934506	Johnstonei_A	4405578	79.98	2948
AS01_indexing48	R173484	-14.608592	126.934506	Johnstonei_A	1777993	27.21	2906
AS01_indexing50	R171488	-15.991667	125.328333	Johnstonei_A	1770111	24.10	2867
AS01_indexing51*	BP01423	-15.949700	124.561100	Johnstonei_A	3019040	46.00	2935
AS01_indexing54	CCM0992	-14.880000	126.358530	Johnstonei_A	1746385	22.12	2802
AS01_indexing55	CCM1046	-15.132800	126.200200	Johnstonei_A	2389877	29.27	2878
AS01_indexing56	R169980	-16.176667	123.639444	Johnstonei_A	2961107	34.28	2921

Table S2.3 Estimates of net evolutionary divergence of sampled exons between species and lineages. Standard error estimates are shown above the diagonal. Values obtained using the concatenated dataset (175,244 positions considered) and MEGA7 (Kumar, Stecher, and Tamura 2016).

	C. triacantha	Johnstonei A	Johnstonei B
C. triacantha		4.00E-05	6.00E-05
Johnstonei A	0.00087		5.00E-05
Johnstonei B	0.00114	0.00123	

	Triacantha A	Triacantha B	Johnstonei A	Johnstonei B
Triacantha A		2.00E-05	3.00E-05	5.00E-05
Triacantha B	0.00026		4.00E-05	6.00E-05
Johnstonei A	0.00093	0.00094		5.00E-05
Johnstonei B	0.00122	0.00119	0.00123	

Table S2.4 Lineage divergence histories summary results for the two tested models. NA and NB correspond to population sizes for lineages A and B within both species. NmAB and NmBA correspond to migration rate measured by number of migrants. Tau in years was calculated with $\theta=4N\mu$, generation time of one year and mutation rate of $9x10^{-10}$ as in Singhal and Moritz (2013). F_{st} obtained from 2D-SFS in ∂a∂i.

		C. johnstonei		*C. triacantha*	
		Strict isolation model	Isolation-with-migration model	Strict isolation model	Isolation-with-migration model
Maximum Log composite likelihood		-621.79	-479.382	-353.10	-349.142
Ancestral theta		289.85	134.002	204.02	233.946
Best-fit parameters	**NA**	1.320	2.551	1.884	1.438
	NB	0.473	0.496	1.329	1.079
	Tau	1.220	4.993	1.197	1.075
	NmAB		0.008		0.086
	NmBA		0.128		0.001
Tau (years)			5.62E+06		2.22E+06
Total sequence length*			66179		62846
F_{st}			0.51		0.37

* Proportion of total sequence length accounting for subsampling one SNP per locus

Table S2.5 Results of the range expansion analysis for three main nuclear lineages and Johnstonei A excluding individuals from Johnstonei A2. Given for each lineage are the latitude and longitude of the inferred origin of the expansion, founder distance (d1), r^2 of the regression of psi vs. difference in distance from origin, and Bonferroni-corrected P-value (P).

	Triacantha A	Triacantha B	Johnstonei A	Johnstonei A without A2
Longitude	125.3673	138.13	126.5877	126.6369
Latitude	-14.77527	-18.6	-14.86535	-14.78822
d1	13.279	85.248	1.919	1.895
r^2	0.258	0.000	0.428	0.446
P	2.208	3269.062	1.92E-18	2.47E-17

Table S2.6 Genetic diversity indices for the EC lineage Johnstonei A with and without identified groups. Given are sample size (n), number of loci studied, average locus size in base pairs (bp), average theta per site from segregating sites (θ_S), average nucleotide diversity (π), and average Tajima's D per locus.

EC lineages	*n*	Number of loci	*bp*	θ_S	π	Tajima's D
Johnstonei A	20	2048	442.20	0.00551	0.00165	-2.245
Without Johnstonei A2	18	2066	481.37	0.00523	0.00167	-2.227
Without Johnstonei A2 and Jungulu and Augustus islands	16	2074	483.14	0.00487	0.00166	-2.211

Table S2.7 ND4 genetic diversity indices for each lineage identified in the mtDNA phylogenetic tree. n, sample size; Hd, haplotype diversity; S, segregating sites; θ_η, theta per site from η (total number of mutations), π, nucleotide diversity, k, average number of nucleotide differences. None of the Tajima's D or Fu's Fs values were significant.

mtDNA Lineages	*n*	Hd	S	θ_η	π	*k*	Tajima's D	Fu's Fs
Triacantha A	46	0.986	143	0.040	0.038	25.279	-0.809	-10.337
Triacantha A - mesic	9	1.000	97	0.062	0.057	38.028	0.338	-0.349
Triacantha A - arid	37	0.979	74	0.020	0.020	13.314	-0.913	-10.992
Triacantha B	116	0.995	213	0.088	0.079	48.421	0.698	-25.602
Johnstonei A1	100	0.991	206	0.072	0.066	44.148	0.368	-15.175
Johnstonei A1 - North	83	0.987	159	0.055	0.051	34.414	0.272	-9.525
Johnstonei A1 - South	17	1.000	117	0.058	0.054	36.132	0.188	-3.256
Johnstonei B	54	0.928	28	0.007	0.007	4.549	-0.849	-6.317
Johnstonei A2	6	0.933	48	0.034	0.044	27.4	1.943	2.519
Johnstonei A3	14	0.912	78	0.032	0.030	20.264	-0.772	2.693

Figure S2.1 mtDNA ND4 maximum likelihood phylogenetic tree of *Carlia triacantha* and *Carlia johnstonei* with clades not collapsed. Sample label includes tissue number, taxon and sampling location.

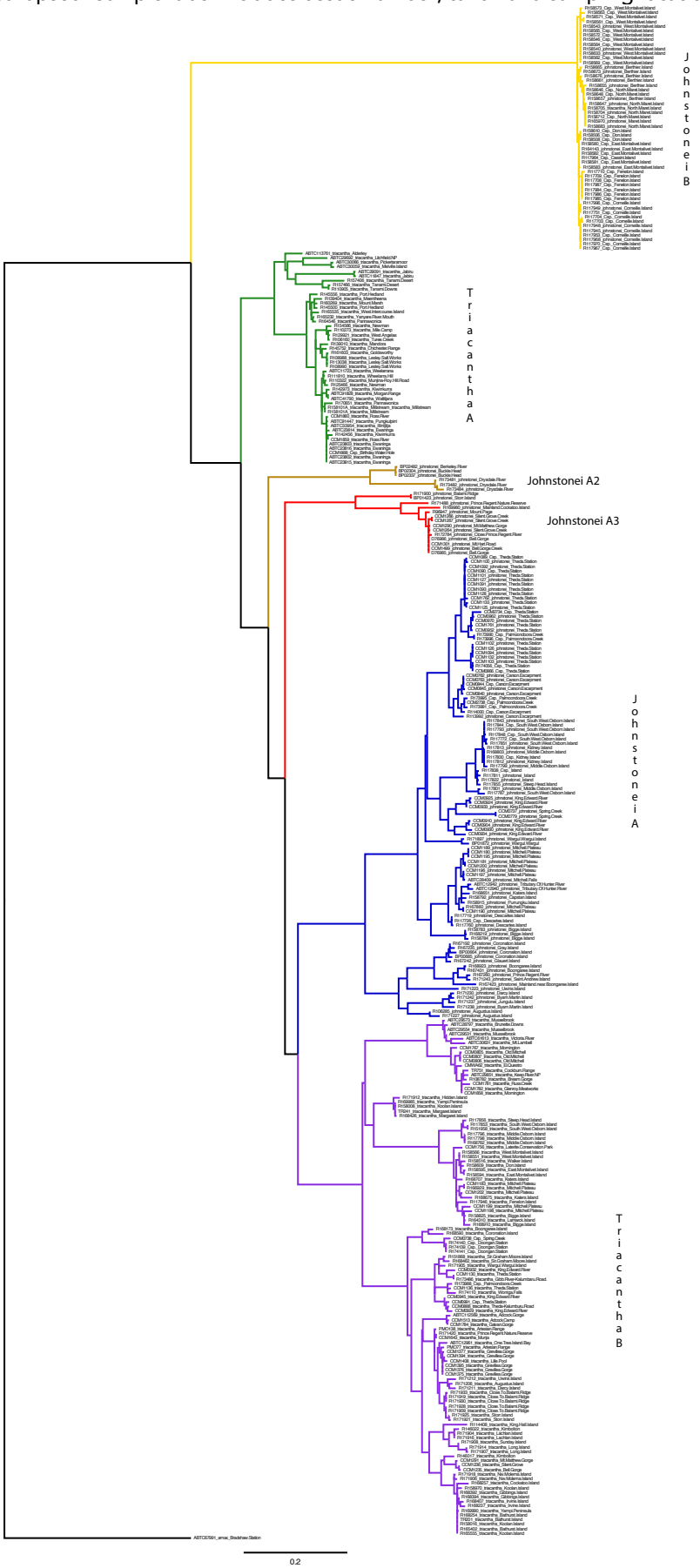


Figure S2.2 Map displaying the cytonuclear discordance in the mesic-arid Triacantha A (green shaded area) split. Also shown is the sampling gap in the center of Northern Territory (NT) where possibly both Triacantha A and Triacantha B (purple shaded area) can overlap.

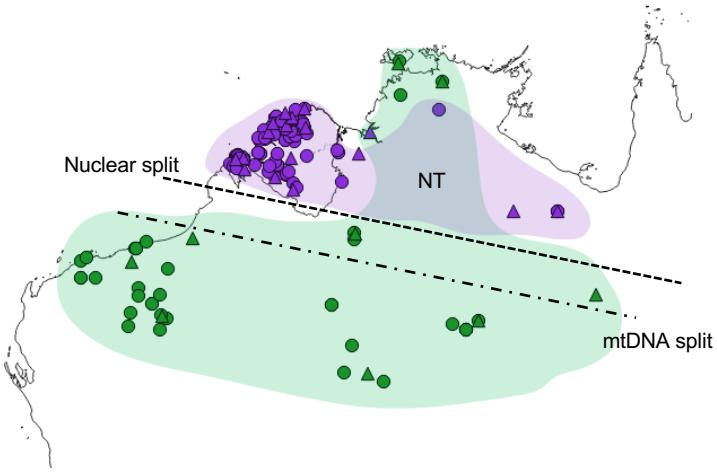


Figure S2.3 Principal coordinate analysis (PCoA) based on Euclidean genetic distances and STRUCTURE results for the three analyzed main lineages. *Samples from Augustus and Jungulu islands.

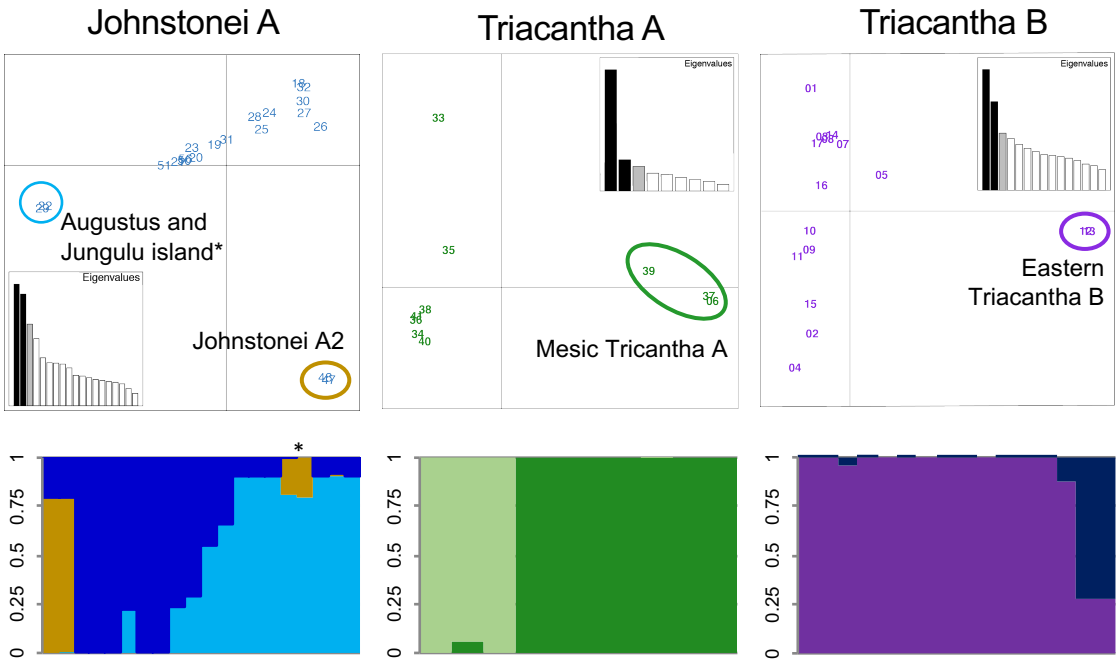


Figure S2.4 Plot of mean annual precipitation with mean annual temperature (Williams et al. 2010), showing the difference in climatic niche breadth between Triacantha B and *C. johnstonei* (blue dots). This considers all known locations for *C. triacantha* (purple dots) in the Kimberley and in NT (white dots).

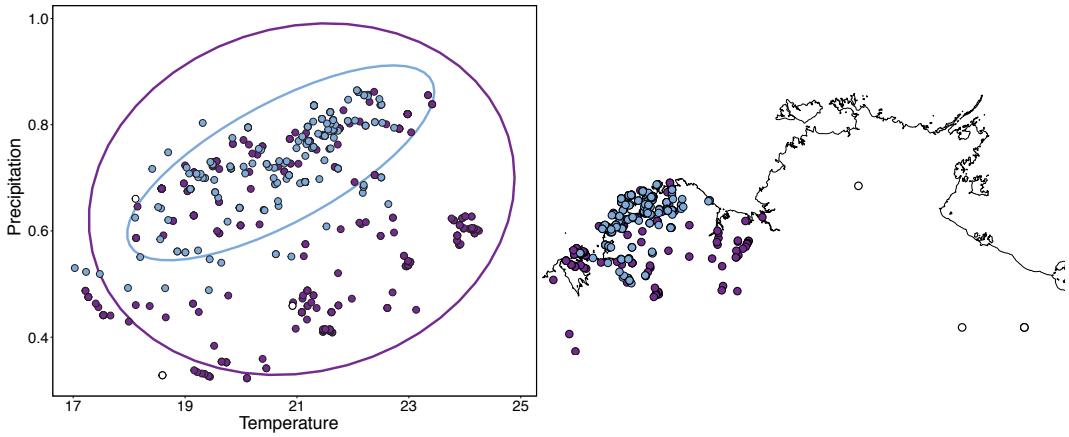
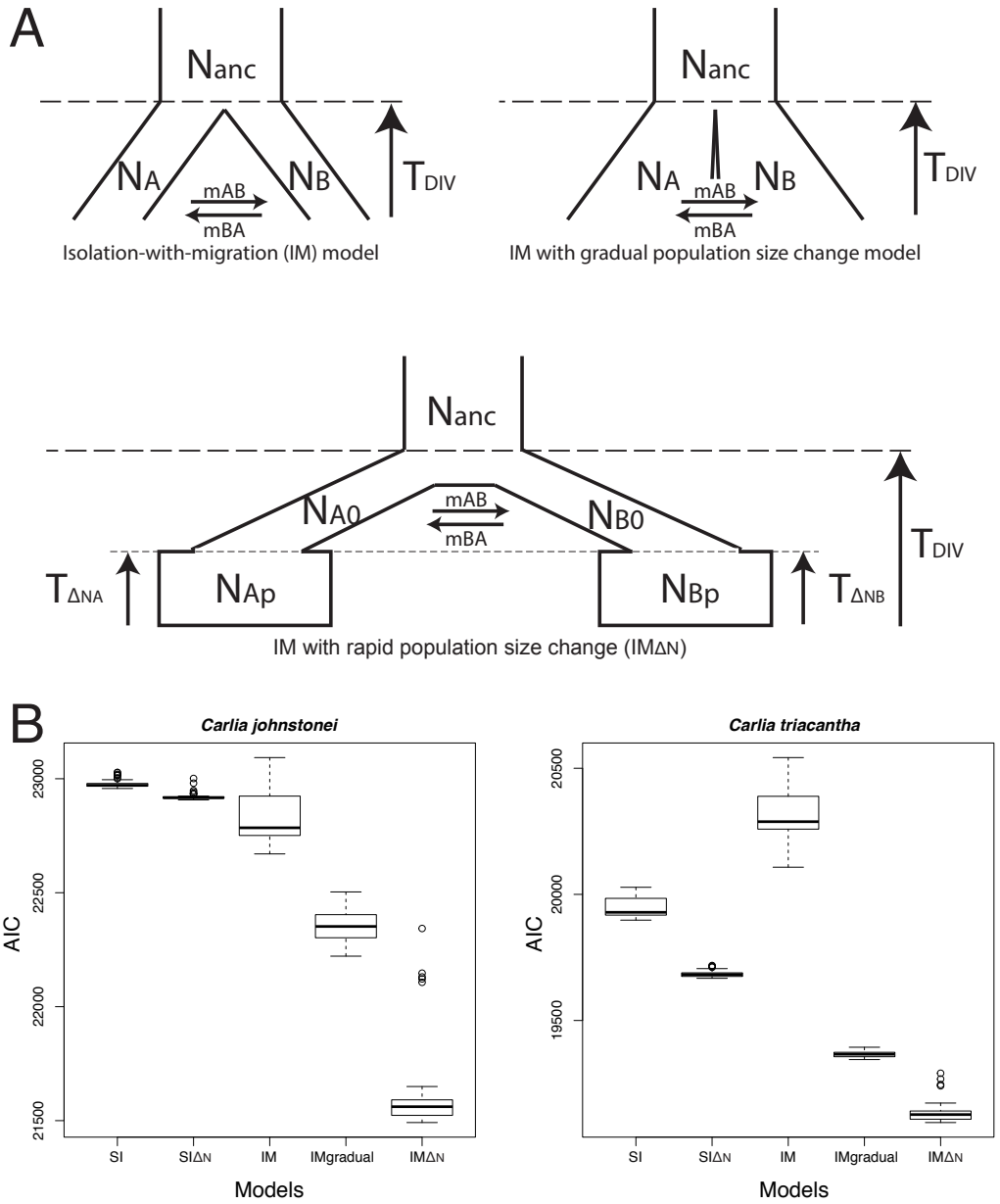


Figure S2.5 (A) Isolation-with-migration models tested with FASTSIMCOAL2; (B) boxplots with AIC values from all replicated runs for the five tested models.



Chapter III

VALIDATION AND DESCRIPTION OF TWO NEW NORTH-WESTERN AUSTRALIAN RAINBOW SKINKS WITH MULTISPECIES COALESCENT METHODS AND MORPHOLOGY

Ana C. Afonso Silva[1,2], Natali Santos[3], Huw A. Ogilvie[1,4] and Craig Moritz[1]

[1]Division of Ecology and Evolution, Research School of Biology and Centre for Biodiversity Analysis, Australian National University, Acton, ACT, Australia.
[2]cE3c - Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.
[3]Universidade Federal do ABC, Santo André, SP, Brazil.
[4]Centre for Computational Evolution, University of Auckland, Auckland, New Zealand.

The original work described in this chapter has been published and should be referenced as:
Afonso Silva AC., Santos N., Ogilvie HA., Moritz C. (2017). Validation and description of two new north-western Australian Rainbow skinks with multispecies coalescent methods and morphology. *PeerJ*, **5**, e3724.

Abstract

While methods for genetic species delimitation have noticeably improved in the last decade, this remains a work in progress. Ideally, model based approaches should be applied and considered jointly with other lines of evidence, primarily morphology and geography, in an integrative taxonomy framework. Deep phylogeographic divergences have been reported for several species of *Carlia* skinks, but only for some eastern taxa have species boundaries been formally tested. The present study does this and revises the taxonomy for two species from northern Australia, *Carlia johnstonei* and *C. triacantha*.

We introduce an approach that is based on the recently published method StarBEAST2, which uses multilocus data to explore the support for alternative species delimitation hypotheses using Bayes Factors (BFD). We apply this method, jointly with two other multispecies coalescent methods, using an extensive (from 2163 exons) data set along with measures of 11 morphological characters. We use this integrated approach to evaluate two new candidate species previously revealed in phylogeographic analyses of rainbow skinks (genus *Carlia*) in Western Australia.

The results based on BFD StarBEAST2, BFD* SNAPP and BPP genetic delimitation, together with morphology, support each of the four recently identified *Carlia* lineages as separate species. The BFD StarBEAST2 approach yielded results highly congruent with those from BFD* SNAPP and BPP. This supports use of the robust multilocus multispecies coalescent StarBEAST2 method for species delimitation, which does not require a priori resolved species or gene trees.

Compared to the situation in *C. triacantha*, morphological divergence was greater between the two lineages within Kimberley endemic *C. johnstonei*, which also had deeper divergent histories. This congruence supports recognition of two species within *C. johnstonei*. Nevertheless, the combined evidence also supports recognition of two taxa within the more widespread *C. triacantha*.

With this work, we describe two new species, *Carlia insularis* sp. nov and *Carlia isostriacantha* sp. nov. in the northwest of Australia. This contributes to increasing recognition that this region of tropical Australia has a rich and unique fauna.

Keywords: Integrative taxonomy, *Carlia*, Species delimitation, Australia, Multispecies coalescent, BFD* StarBeast2

3.1 Introduction

Cryptic species – when two or more distinct species are inaccurately classified under one species name (Bickford et al., 2007) - present great challenges for taxonomy and species delimitation due to the desirability of validating candidate species using multiple lines of evidence (Fujita et al., 2012). But for biodiversity assessment and conservation reasons the need to properly describe species diversity is greater than ever (Bickford et al., 2007). In the same way, there is a concern that molecular data may promote taxonomic inflation by 'over splitting' divergent populations into candidate species (Isaac, Mallet & Mace, 2004; Hedin, Carlson & Coyle, 2015). The creation of more reliable and robust species delimitation approaches in the last decade has attempted to address this concern (Rannala, 2015).

To more robustly infer species boundaries, the use of integrative taxonomy is increasingly common (Bickford et al., 2007; Padial et al., 2010). The objective of this approach is to corroborate taxonomic validity with independent, distinct types of evidence. Given deep genetic divergence, fixed morphological differences are not necessary to diagnose species boundaries since speciation itself does not require phenotypic characters to evolve at the same rate as the genome (Leaché & Fujita, 2010). Therefore, in taxa with inherently conservative morphology, it may be that the primary evidence for distinct species will come from genetic data.

Species delimitation consists of two potentially complementary approaches: discovery methods that do not require a priori assignment of samples before analysis, and validation methods that test hypotheses based on samples already assigned to candidate species (Ence & Carstens, 2011). When candidate lineages are already identified, validation approaches are more robust because they explicitly model the process of lineage diversification (Carstens et al., 2013). This is especially so when there is a substantial number of informative genes, independent of those used to suggest candidate taxa. Model-based multilocus approaches that use the multispecies coalescent (MSC) are advantageous because they account for coalescent processes when estimating phylogenetic relationships (Edwards et al., 2016). And for species delimitation, objective and transparent model-based approaches are relevant, because they have the potential to reduce investigator-driven biases (Fujita et al., 2012). These methods can consider gene tree incongruence due to incomplete lineage sorting, variation in molecular sequences and variation in demographic parameters (Leaché & Fujita, 2010). With this in mind, Carstens et al. (2013) recommend the best approach for species delimitation is to use multiple methods. Further, Rannala (2015) suggests that this should only be done when methods have algorithmically similar assumptions. However, we also note that MSC methods can over split – revealing high structured populations (or ephemeral species; Rosenblum et al., 2012) – rather than long isolated species, depending on the nature of the speciation process (Sukumaran & Knowles, 2017). Hence, species delimitation will always be more secure when taxa delimited using genetic methods are somehow corroborated by alternative sources of data (Oliver, Keogh & Moritz, 2015).

Previous work by Afonso Silva et al. (2017), which focused on understanding how phylogeographic structure and history differs between a climatic generalist and specialist, found two deeply divergent lineages within each of *Carlia johnstonei* Storr, 1974 and *C. triacantha* Mitchell, 1953 (Fig. 3.1). These sister taxa (Dolman & Hugall, 2008) have contrasting distributions, with the former being endemic to the Kimberley and the latter being widespread across northern Australia. The lineages within *C. johnstonei* are likely allopatric, with the nominal lineage (Johnstonei A) being found across the north and western Kimberley and the newly identified lineage (Johnstonei B) being endemic to islands off the coast of the northwest Kimberley (Fig. 3.1). Conversely, the two lineages of *C. triacantha* likely overlap geographically, with the nominal lineage being widespread across north and central Australia (Triacantha A) and the newly identified lineage (Triacantha B) found within the Kimberley and scattered locations in the central Northern Territory (Fig. 3.1).

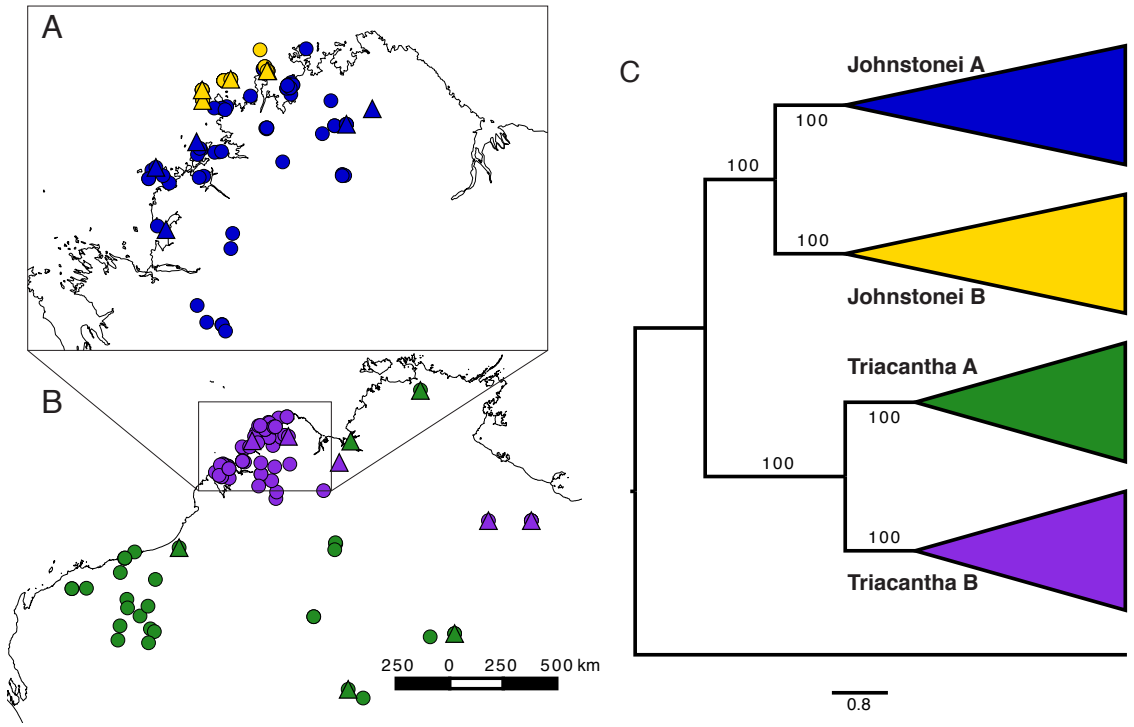


Figure 3.1 Distribution map with used genetic samples and measured specimens for *C. johnstonei* (A) and for *C. triacantha* (B) lineages, and lineages relationships (C) as in Afonso Silva et al. (2017). Triangles correspond to the genetic samples used in this study while circles correspond to specimens measured. Blue Johnstonei A, yellow Johnstonei B, green Triacantha A and purple Triacantha B. Tree obtained with 20 representative samples in ASTRAL and respective lineage bootstrap.

Species of *Carlia* from the Australian tropics generally have deep phylogeographic structure for both mtDNA and large numbers of exons (e.g. Potter et al., 2016), and, where contact zones have been examined in detail, there is evidence of strong reproductive isolation between the more deeply divergent (but phenotypically cryptic) lineages (Phillips, Baird & Moritz, 2004; Singhal & Moritz, 2013). However, recent species delimitation and taxonomic revisions have focussed more on *Carlia* from the eastern woodlands and placed a greater emphasis on morphology (e.g. Hoskin & Couper, 2012; Hoskin, 2014). There is a

need to re-examine the systematics of northern Australian *Carlia*, and here we have the opportunity to exploit large multilocus datasets (Bragg et al., 2015) to do that integrated with morphology. This is particularly relevant for Kimberley biodiversity, since there have been recent efforts to discover and describe new species (Köhler, 2011; Oliver et al., 2014, 2016; Andersen et al., 2014; Ellis, 2016) in this still relatively unknown and remote region in the northwestern of Australia.

We reanalyse the extensive multilocus data used in Afonso Silva et al. (2017) using robust species delimitation methods, together with morphological analysis to validate species hypotheses. Following Rannala (2015), we use three algorithmically similar methods to validate potential new species. We apply BPP (Yang & Rannala, 2014) and two approaches using Bayes Factors to test species hypotheses: a SNP based approach, BFD* (Leaché et al., 2014) using SNAPP (Bryant et al., 2012) and a sequence-based approach, BFD with the recently developed StarBEAST2 method (Ogilvie, Bouckaert & Drummond, 2017). We consider three potential species hypotheses: (i) only the two currently defined species are separated; (ii) a three-species hypothesis - two species corresponding to the two more deeply divergent lineages of *C. johnstonei*, but collapsing the less divergent lineages within *C. triacantha*; and (iii) a four-species hypothesis - all four lineages correspond to different species. Using an integrative taxonomic approach, we present and analyse morphological data to test for congruent differences between all identified genetic lineages. Considering all lines of evidence, we then formally describe the new species and identify diagnostic traits, for both morphology and gene sequences. Genetic diagnostic traits include SNPs from available mtDNA ND4 gene sequences (Afonso Silva et al. 2017), following Renner's (2016) suggestion to provide simple genetic diagnostics, particularly for morphologically similar species groups.

3.2 Materials and Methods

We used exon capture data to perform validation analyses and sequences of the mtDNA ND4 gene to identify diagnostic SNPs, and, also measured, analysed and identified diagnostic morphological traits. We obtained sequences for the genetic data from Afonso Silva et al. (2017) (Dryad Digital Repository http://dx.doi.org/10.5061/dryad.jj1t). These included mtDNA sequence data of 101 *C. johnstonei* and 99 *C. triacantha* throughout both species' distribution, for which we had specimens to do morphological analysis (Table S3.1, Fig. S3.1).

See Afonso Silva et al. (2017) for more detail about how the exon capture data was obtained. In summary, the data was retrieved from a custom set of loci designed from transcriptomes of *Carlia* and a couple of related genera (Bragg et al., 2015). After similar processing to Bragg et al. (2015), the final dataset contained a total of 51 samples with average of 40x coverage and approximately 2800 loci per sample. For the validation analyses, we retrieved data from the 20 geographically dispersed samples as used for species tree

estimation in Afonso Silva et al. (2017) (Fig. 3.1, Table S3.1). These correspond to five individuals for each of the four lineages previously identified in Afonso Silva et al. (2017) (Fig. 3.1, Table S3.1), using the same *C. amax* samples as an outgroup (from Potter et al., 2016) that were used in that study. For these analyses, we required aligned haplotype sequences, for which we employed GATK (v3.3, McKenna et al., 2010) which was also used to identify heterozygous sites and mask sites with a low-quality genotype call (GQ<20). Here, we generated phased haplotypes using the individual overlapping sequencing reads to phase heterozygous sites within target loci and then used one haplotype per sample in later analyses.

We then used the EAPhy pipeline (Blom 2015 v.1.2; github.com/MozesBlom/EAPhy) to realign, filter and export alignments with complete data into NEXUS and PHYLIP format, as well as two sets of SNPs in FASTA format (using 0.2 as maximum proportion of Ns for each site, one SNP chosen randomly per gene and excluding singletons).

3.2.1 Genetic species validation

We applied three multispecies coalescent validation approaches to investigate species boundaries: Bayesian Phylogenetics and Phylogeography (BPP v3.3; Yang & Rannala, 2014), BFD (Bayes factor delimitation; Grummer et al. 2014) StarBEAST2 using multilocus data (Ogilvie, Bouckaert & Drummond, 2017), and BFD* SNAPP using SNP data (Leaché et al., 2014).

For the BPP analysis, we randomly selected two exon sets (to avoid unforeseen biases), each with 100 loci of between 250 bp and 1000 bp, to check for consistent results. The MSC assumes no recombination within loci, and free recombination among loci (Degnan et al., 2009). We are confident of satisfying the latter condition, as our exons are all derived from different genes (Bragg et al., 2015). Lanier & Knowles (2012) showed that intra-locus recombination had little effect in species-tree estimates under the MSC; however Potter et al. (2016) found that it can affect species delimitation. Hence, to further evaluate this effect, we used the program IMgc (Woerner, Cox & Hammer, 2007) to extract optimal recombination-filtered blocks (no four-gamete violations) and repeated BPP analysis for comparison. We performed joint Bayesian species delimitation and species tree estimation (method A11, Yang, 2015). This method uses the multispecies coalescent model to compare different models of species delimitation and species phylogeny in a Bayesian framework, accounting for incomplete lineage sorting due to ancestral polymorphism and gene tree species tree conflicts (Yang & Rannala, 2010, 2014; Rannala & Yang, 2013). Ancestral population size parameters (theta) were set to gamma prior G(2, 1000), with mean 2/1000 = 0.002 and the divergence time at the root of the species tree (tau) was assigned to G(2, 2000), while the other divergence time parameters were assigned to the Dirichlet prior (Yang & Rannala, 2010: equation 2). Preliminary analyses run using different combination of gamma priors, as suggested in Yang (2015), produced similar results, suggesting that our results are

robust to the priors used (data not shown). The phylogeny obtained in Afonso Silva et al. (2017) was used as a starting tree and all columns in the alignment were used in the likelihood calculation. Each exon set analysis was independently run twice to confirm consistency between runs, with a burn-in of 50,000 and a sampling frequency of five iterations for a total of 500,000 generations.

Bayes factor delimitation (BFD; Grummer, Bryson & Reeder, 2014) is an approach that compares the marginal likelihoods of competing species delimitation hypotheses using Bayes factors. To apply this approach, we ran two MSC methods to test our three potential hypotheses using *C. amax* as an outgroup: (i) a scenario with two species (*C. johnstonei* and *C. triacantha*), (ii) a scenario with three species (lineages Johnstonei A, Johnstonei B and *C. triacantha*) and (iii) a scenario with four species (with both lineages from *C. triacantha* and *C. johnstonei* as separate species).

StarBEAST2 v0.13.5 is a recently released sequence-based approach that reconstructs species trees with more flexibility than BPP (Ogilvie, Bouckaert & Drummond, 2017), and so provides an alternative MSC method to investigate species delimitation with Bayes factors (BFD). To verify consistency, we randomly selected another two sets of exons, each with 20 loci between 250 and 1000 bp. We then used jModelTest v2.1.10 (Guindon, Gascuel & Rannala, 2003; Darriba et al., 2012) to calculate nucleotide substitution model likelihood scores for each locus and to estimate optimal model using BIC (Supplemental Table S3.2). All BFD StarBEAST2 analyses were performed using a strict clock model, for 100,000,000 generations, with data sampled every 10,000 generations, the first 10% of each run was discarded as burn-in and priors as in Table S3.3. For each analysis, two BFD StarBEAST2 replicates were conducted to ensure convergence and assessed using ESS values with Tracer v1.6 (Rambaut et al., 2015). We used stepping-stone sampling (Leaché et al., 2014) to determine the marginal likelihoods of four, three and two species (plus outgroup). All stepping-stone analyses used 16 steps with a beta distribution α parameter of 0.1 to optimise the power posterior discretization (Xie et al., 2010). The resulting marginal likelihoods were then used to compute Bayes factors (Kass & Raftery, 1995), quantifying the support for each species delimitation hypothesis against all others under consideration. The final tree was obtained by combining posterior replicates with LogCombiner (Drummond & Rambaut, 2007) and summarised using maximum clade credibility trees, after exclusion of 10% burn-in, with TreeAnnotator v1.7.2 (Drummond & Rambaut, 2007).

To use an approach that considers evidence from all available loci, we selected two independent SNP sets by sampling one SNP at random from each locus out of 2,163 total available loci and estimated species trees for each scenario using SNAPP (Bryant et al., 2012). We ran all analysis for 500,000 generations sampling every 500, with two replicates to ensure convergence and priors as in Table S3.3. After assessing convergence between runs and exon sets we proceeded to Bayes factor delimitation as described previously.

3.2.2 Morphological data collection

We analysed 200 specimens from the Museum and Art Gallery of the Northern Territory (MAGNT), Museum Victoria (MV), South Australian Museum (SAM), Western Australian Museum (WAM) and recently-collected specimens held at the Australian National University (with ANU ethical approval number A2012/14) (Fig. 3.1, Table S3.1). All analysed specimens were also sequenced for the mtDNA ND4 gene in Afonso Silva et al. (2017) (Fig. S3.1), with a total of 66 examined specimens for Johnstonei A, 35 for Johnstonei B, 31 for Triacantha A and 68 for Triacantha B.

We examined five morphometric characters taken to the nearest 0.1 mm with Mitutoyo electronic callipers: snout-vent length (SVL), axilla-groin length (AGL), head length (HL) measured from anterior edge of tympanum to snout, head width (HW) measured at widest point of the head, and head depth (HD) measured at parietal scales. In order to minimize error, we used a dissecting microscope Leica MZ8 (equipped with camera Leica MC120 HD) for which forelimb (FLL) and hindlimb length (HLL) were measured through photographs using ImageJ (Abràmoff, Magalhães & Ram, 2004) (as in Fig. S3.3); as well as four additional smaller features: nasals separation (NS), ear aperture length (EAL), palpebral disc length (PDL) and eye to ear distance (EED) (as explained in Fig. S3.3).

We also assessed seven meristic characters using photographs: supralabials, infralabials, supraciliaries, lamellae under the 4th toe (from claw sheath to junction of 3rd and 4th toes), lamellae under the 3rd finger (from claw sheath to the junction of the 2nd and 3rd fingers), the mode of number of keels across the mid-dorsal line scales and the ear lobule numbers. These traits were counted as suggested by Cogger (2014) and similarly to Hoskin & Couper (2012). Measurements and scales were generally analysed from the left side of the specimen, unless prevented by damage or poor preservation. All described measurements were collected in millimetres (mm).

For the ensuing species descriptions, we also measured the tail length and the distance between prefrontals if not in contact, but these traits were not used in the morphological analysis due to high level of missing data. For the designated holotypes, we additionally counted the number of midbody scale rows, vertebral (from the occiput to the edge of the hind limb along the mid dorsal line) and ventral scales (from mental scale to the edge of cloaca).

3.2.3 Morphological analyses

We investigated the relationship of each linear measurement with size (per mtDNA lineage), by plotting each variable against SVL and by comparing box plots of raw and size corrected measurements. After removing samples with missing data, all measurements were log-transformed to reduce their variance allowing a more conservative assessment of differences between mtDNA lineages. We then extracted size-corrected residuals from regressions between SVL and each measurement as a size-corrected log-transformed

dataset. We investigated normality and heteroscedasticity after variable correction using density plots, Shapiro-Wilk test and Levene's test. Multivariate normality was assessed with the Henze-Zirkler's Multivariate Normality Test in the MVN package (Korkmaz, Goksuluk & Zararsiz, 2014).

In order to assess the morphometric distinctiveness of these lineages, we conducted Principal Component analyses on the log-transformed and on the size-corrected log-transformed (excluding SVL) datasets for each species. We used the prcomp function (stats package) with all measurement variables centred and plotted principal component 1 (PC1) against PC2, with a 75% confidence ellipse probability threshold (ggplot2 package, Wickham 2016).

To statistically evaluate whether the lineages are significantly different and which variables are contributing to this, we analysed log-transformed and size-corrected log-transformed measurements with a MANOVA, and confirmed the significance of non-normal variables with the non-parametric Wilcoxon test (stats package). Relevant meristic data was analysed independently with a generalized linear modelling with a Poisson distribution (stats package) since these are count data and not continuous variables.

Using the statistically significant measurement variables from the MANOVA, we tested the accuracy in predicting assignment of lineage by applying a linear discriminant analysis (LDA) with jackknife cross-validation implemented in the package MASS (Ripley et al., 2013). Due to the presence of non-normal variables, we also applied a Random Forest (RF) analysis using the package randomForest (Liaw & Wiener, 2002).

We investigated the effect of possible outliers in the data by calculating, for each of the variables, interquartile range scores (function scores in outliers package, Komsta, 2011) to identify samples with outliers and then perform a MANOVA with this dataset. Removing outliers decreases 14% and 6% of analysed specimens for *C. johnstonei* and *C. triacantha*, respectively. Since some of these outliers could represent expected phenotypic variation across these species distribution and the overall results were similar, we present the analyses with all individuals.

To account for the insufficient information on sex, we performed a linear model containing sex and mtDNA lineage, using the available sexed individuals, which showed no difference in SVL between males and females in either *C. johnstonei* or *C. triacantha*. This suggests sex differences cannot explain our observed results, so we also present the analyses with all individuals.

We performed all analyses in R v.3.3.1 (R Core Team, 2016) and all the data, input files, code and morphological results are available at:

https://dx.doi.org/10.6084/m9.Figshare.4621963.

3.2.4 Molecular diagnostics

Following the recommendation of Renner (2016) , we visually identified diagnostic

SNPs within the ND4 mtDNA gene using all Afonso Silva et al. (2017) sequences with Genbank accessions codes MF083173-MF083508 in Geneious v.7.1.9 (http://www.geneious.com, Kearse et al., 2012). Using as a reference an available skink mitogenome from Scincella vandenburghi (Park et al., 2016), we selected the available diagnostic SNPs per lineage within each species, where the nucleotide difference would correspond to an amino acid substitution.

The electronic version of this article in Portable Document Format (PDF) will represent a published work according to the International Commission on Zoological Nomenclature (ICZN), and hence the new names contained in the electronic version are effectively published under that Code from the electronic edition alone. This published work and the nomenclatural acts it contains have been registered in ZooBank, the online registration system for the ICZN. The ZooBank LSIDs (Life Science Identifiers) can be resolved and the associated information viewed through any standard web browser by appending the LSID to the prefix http://zoobank.org/. The LSID for this publication is: urn:lsid:zoobank.org:pub:A7B29F16-079F-48BA-B4BE-3EC9A3D80D34. The online version of this work is archived and available from the following digital repositories: PeerJ, PubMed Central and CLOCKSS.

3.3 Results

3.3.1 MSC Species delimitation

All three MSC approaches assigned more support to the four-species hypothesis than either the two- or three-species hypotheses (Table 3.1).

Both BPP analyses, each with independent drawn sets of genes, yielded the same species tree (Fig. 3.2) and a posterior probability (PP) equal to 1 for five delimited species (all four lineages plus the outgroup). The analyses processed with IMgc to exclude blocks with no four-gamete violations from within alignments, returned similar results with PP = 1 for four lineages plus outgroup. However, while topology for the original datasets was as expected by 99% of the models (Fig. 3.2), for each gene set without recombining blocks only 64% and 85% of the models supported the same topology.

For both BFD StarBEAST2 and BFD* SNAPP, Bayes Factors (BF) were obtained by subtracting the two-species hypothesis from both the three-species hypothesis as well the four-species hypothesis, and multiplying the difference of marginal likelihoods by a factor of two.

The BFs for both the BFD StarBEAST2 and BFD* SNAPP analyses were > 10 for the four-species hypothesis relative to the two- or three-species hypotheses (Table 3.1), which corresponds to decisive evidence for this model (Kass & Raftery, 1995). The marginal likelihood results were of similar magnitude across the two gene datasets for BFD StarBeast2 and across the two SNP datasets for BFD* SNAPP (Table 3.1, Fig. S3.2), although BFs were

much higher for the latter.

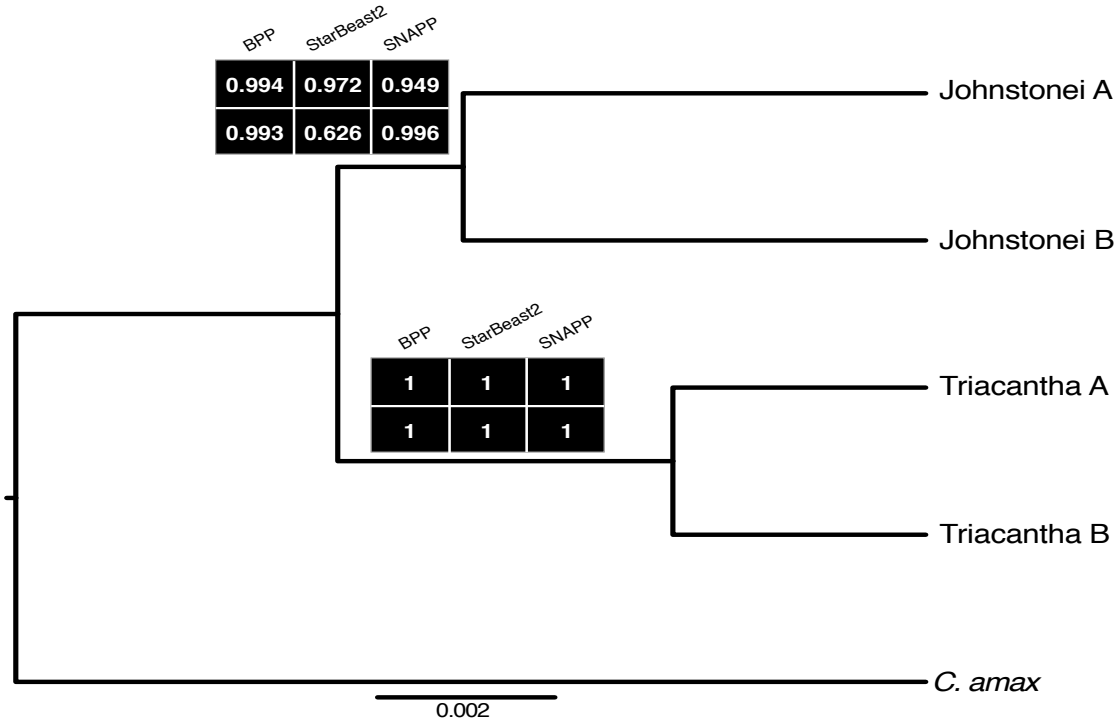


Figure 3.2 Species tree with topology from BFD StarBeast2 gene set1 presenting node posterior probabilities for the two sets of data used for all three MSC methods.

The species tree topology with the main lineages was assessed in Afonso Silva et al. (2017) using the ASTRAL summary species tree method (Fig. 3.1), but here species trees were also estimated by BPP, StarBEAST2 and SNAPP. StarBEAST2 and SNAPP all returned majority support for the ASTRAL topology. For gene sets 1 and 2, StarBEAST2 support for the ASTRAL topology was 97% and 63%, respectively. Support was higher using SNAPP at 95% and >99% for SNP set 1 and 2, respectively.

Table 3.1 Species delimitation support. For BPP support is in posterior probabilities while for BFD StarBeast2 and BFD* SNAPP is based in Bayes Factors calculated using the two species model as the null model (two species support by comparing with the four species model).

	BPP gene set1	BPP gene set2	StarBeast2 gene set1	StarBeast2 gene set2	SNAPP SNP set1	SNAPP SNP set2
Two species	0	0	-318.10	-274.04	-4517.80	-4443.35
Three species	0	0	223.35	203.79	3526.41	3370.49
Four species	1	1	318.10	274.04	4517.80	4443.35

3.3.2 Morphological analysis

The morphological measurements suggest that snout-vent length (SVL) is an important differentiating trait between candidate species within each of *C. johnstonei* and *C. triacantha*

(Fig. 3, S4). Thus, further analyses were conducted also using size-corrected log-transformed variables (Fig. S5), so we could assess if the lineages were statistically different after accounting for SVL differences. For multivariate analyses, individuals with missing data were removed and after size correction some variables were still not normal (Table S3.4, S3.5), but were multivariate normal for both *C. johnstonei* (log-transformed HZ p-value = 0.056, size corrected HZ p-value = 0.121) and *C. triacantha* (log-transformed HZ p-value = 0.104, size corrected HZ p-value = 0.272).

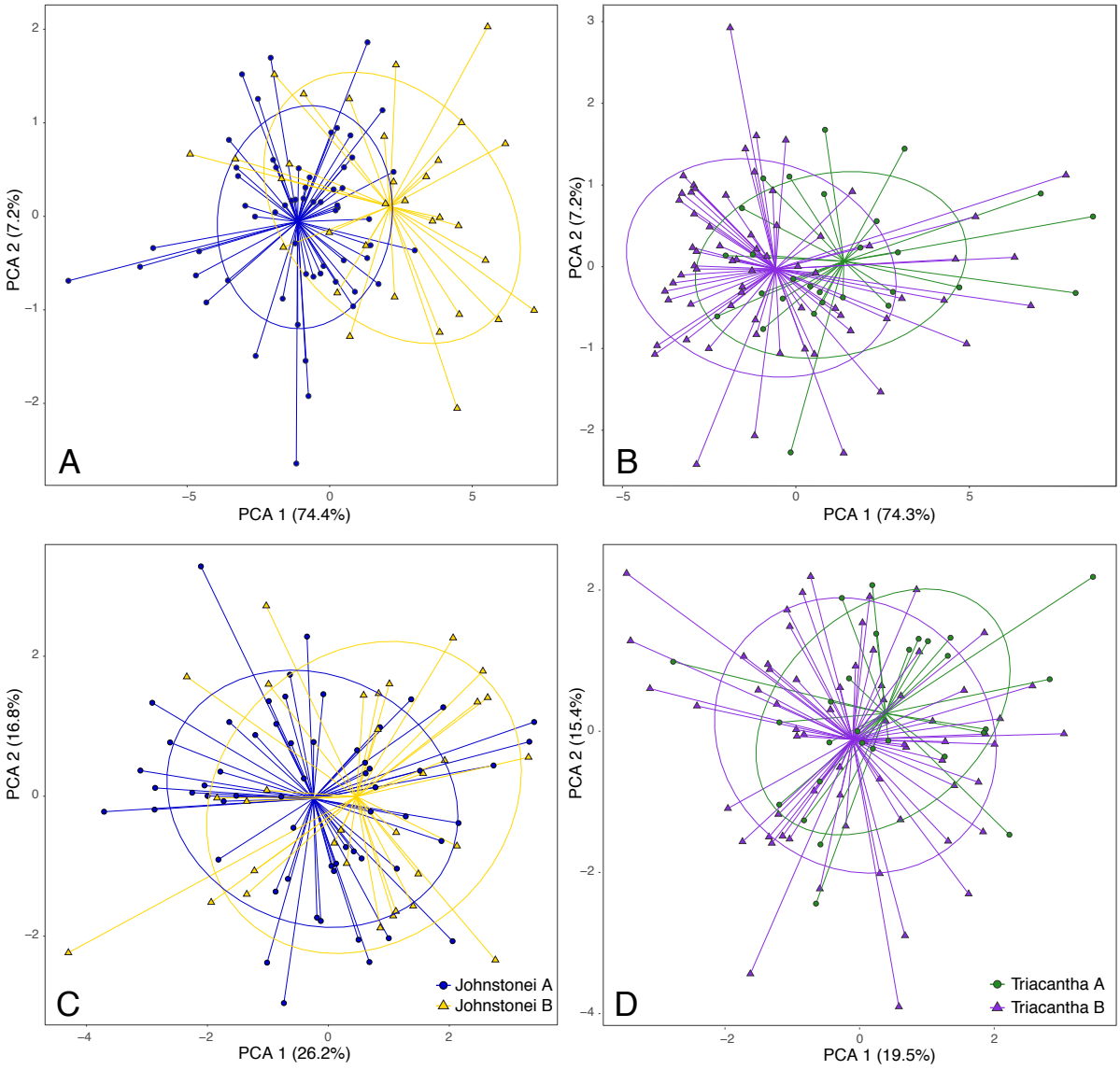


Figure 3.3 PCA with log transformed (A, B) and size corrected (C, D) morphological measurements for *C. johnstonei* and for *C. triacantha* with colours by mtDNA lineage.

In the PCA results for *C. johnstonei* with only log-tranformed data (including SVL), the first axis (PC1) explained 74.4% of the total variation with all variables loading uniformly (and hence size-related) and the second axis (PC2) only explained 7.2% of variation (Fig. 3.3A, S3.6A). By contrast, in the PCA with the size corrected dataset (and excluding SVL), PC1 explains 26.2% and PC2 16.8% of the variation (Fig. 3.3C, S3.6C). The log-transformed PCA shows more evidence of clustering by lineage than does the size-corrected PCA. Together

these observations point to a high similarity in shape, relative to divergence in body size. For *C. triacantha*, similar results were obtained (Fig. 3.3B, S3.6B). The proportions of variance explained for log transformed analysis were PC1 = 74.3% and PC2 = 7.2%; whereas, for the size corrected analysis, PC1 = 19.5% and PC2 = 15.4% (Fig. 3.3D, S3.6D).

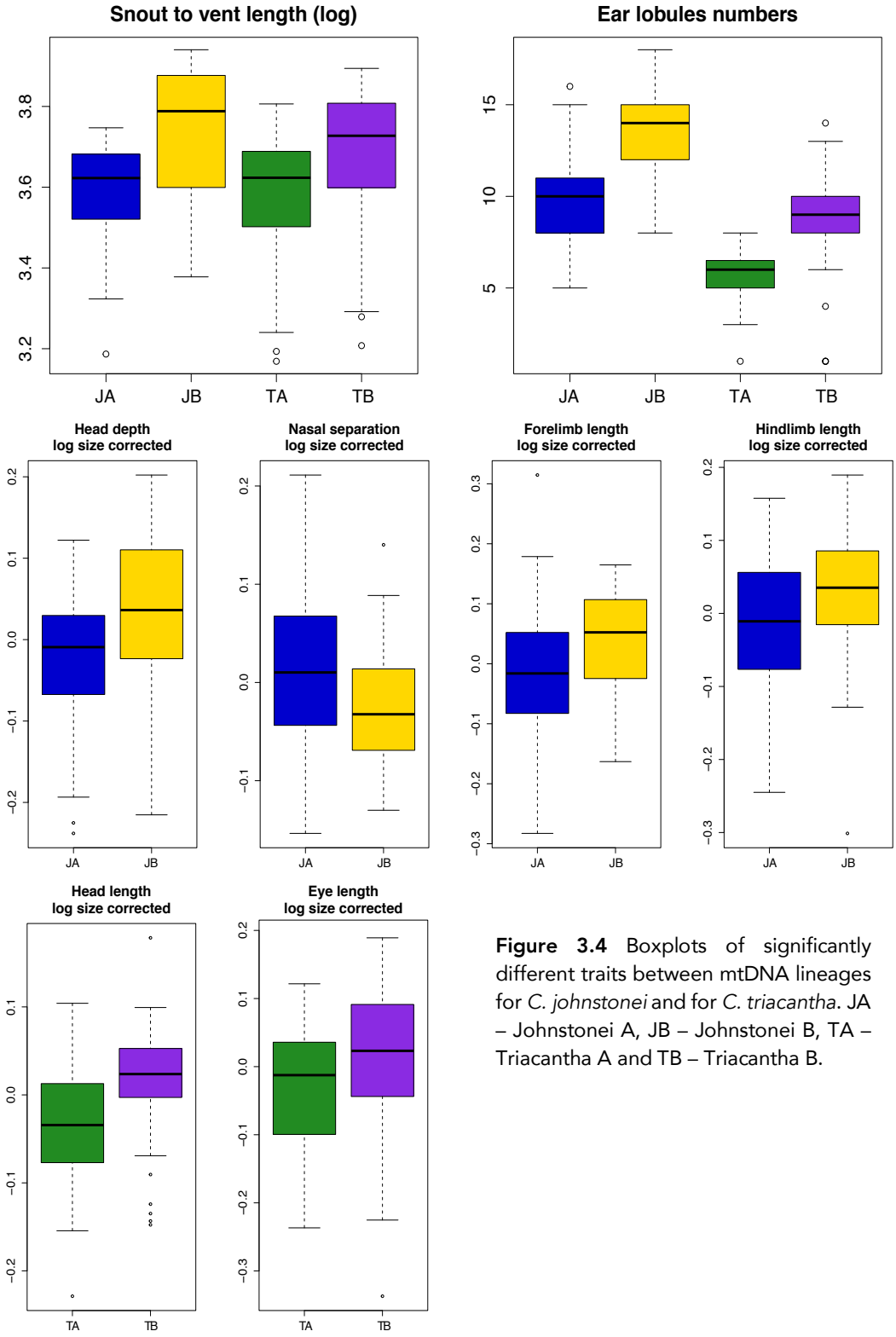


Figure 3.4 Boxplots of significantly different traits between mtDNA lineages for *C. johnstonei* and for *C. triacantha*. JA – Johnstonei A, JB – Johnstonei B, TA – Triacantha A and TB – Triacantha B.

Using MANOVA, we assessed whether morphological measurements differences between lineages were significant (for more detail see Tables S3.4, S3.5). For both species, the MANOVA confirmed that size (logSVL) differs between lineages in each species (p-value = $1.05x10^{-6}$ in *C. johnstonei*; p-value = $6.96x10^{-3}$ in *C. triacantha*). For size-corrected data, head depth (p-value = $1.36x10^{-3}$), nasal separation (p-value = $9.02x10^{-3}$), forelimb (p-value = $7.89x10^{-3}$), and hindlimb (p-value = $2.55x10^{-2}$) are important traits in distinguishing Johnstonei A from Johnstonei B; and head length (p-value = $3.30x10^{-4}$) and ear to eye distance (p-value = $2.73x10^{-2}$) for distinguishing Triacantha A from Triacantha B (Fig. 3.4). The significant non-normal variables within *C. triacantha* were confirmed with significant non-parametric test (Table S3.5).

The analysis of meristic data was based on three relevant characters (number of ear lobules, lamellae number under the 3rd finger and under the 4th toe) due to little or no variation in the other traits. Each of the three analysed characters was significantly different between Johnstonei A and B, but only ear lobule number showed a significant difference between Triacantha A and B (Fig. 3.4, Table S3.6).

The prediction capacity of significant morphological data was investigated with a linear discriminant analysis (LDA) and a Random Forest analysis (RF). Jackknife results provided 85.87% accuracy for differentiating *C. johnstonei* lineages based on log-transformed morphological measurements (to include SVL as a variable) and 72.34% for *C. triacantha* lineages. While the accuracy estimated with a RF analysis was 81.52% for *C. johnstonei* and 68.09% for *C. triacantha*. The summary of each measured trait can be found in supplementary Table S3.7.

Table 3.2 ND4 mtDNA diagnostic SNPs for each lineage. The position of each SNP is aligned with *Scincella vandenburghi* mitochondrial genome (Park et al. 2016). For each nucleotide position is also presented the correspondent amino acid substitution. Grey background refers to which species the SNP is diagnostic for.

	10851		10864		10992		11115		11218		11365		11413	
C. johnstonei	T	Ser	A	Tyr	A	Thr	A	Met	C	Thr	A	Asn	T*	Ile*
C. insularis sp. nov.	A	Thr	G	Cys	A	Thr	A	Met	T	Ile	G	Ser	T	Ile
C. triacantha	T	Ser	A	Tyr	G	Ala	C	Leu	C	Thr	G	Ser	C	Thr
C. isostriacantha sp. nov.	T	Ser	A	Tyr	A	Thr	A	Met	C	Thr	G	Ser	T	Ile

*Substitution is not diagnostic for a few individuals

3.3.3 Taxonomic assessment and species description

Considering the congruence across multiple genetic delimitation methods and of these with significant morphological divergence among lineages, we provide sufficient evidence for four species, two species within *Carlia johnstonei* and two species within *C. triacantha*.

.

Within *C. johnstonei*, Johnstonei A is the nominal *C. johnstonei* species based on a holotype from the Mitchell Plateau, a region in which extensive sampling has shown that only

Johnstonei A occurs. For *C. triacantha* the holotype specimen is from Adelaide River, Northern Territory, a site close (~ 15 km) to Triacantha A samples from Litchfield National Park (NTM R22162) – hence we suggest that Triacantha A should retain the species name. Accordingly, we here describe two new species - Johnstonei B as *Carlia insularis* sp. nov. and Triacantha B as *Carlia isostriacantha* sp. nov. In the following we provide diagnoses for the four species. Simple genetic diagnostics (mtDNA diagnostic SNPs; Table 3.2) are robust. For morphology alone, single traits mostly have overlapping ranges, but in combination with each other and geography, should be practical in the field.

Carlia johnstonei Storr, 1974 Records of the Western Australian Museum, Vol. 3, 151-165
Rough brown rainbow-skink
Holotype. WAM R43170, from Mitchell Plateau, Western Australia, in -14.866667 125.833333.
Diagnosis. Dark blackish *Carlia* morphologically distinguishable from geographically overlapping species with a combination of mid-dorsal scales bicarinate (two keels), more numerous supraciliares (usually 7 vs. 6 – *C. amax*, *C. munda*, *C. rufilatus*, *C. isostriacantha* sp. nov., or 5 – *C. gracilis*), larger ear aperture with numerous sharply pointed lobules (mean of 10 lobules), but typically less than in *C. insularis* sp. nov. (mean of 13 lobules). Further distinguished from the latter by smaller body size (mean 36.39 mm vs. 41.83 mm), reduced head depth (mean 3.59 mm vs. 4.48 mm), shorter limbs (forelimbs 9.51 mm vs. 11.45 and hindlimbs 14.82 mm vs. 17.77 mm) and less lamellae under longest finger (mean 16.75 mm vs. 19.69 mm) and toe (mean 22.83 mm vs. 26.31 mm).
Description. Snout-vent length (mm): 21.84 - 43.49 (N = 66, mean 36.39). Tail: 27.1 – 61.28 (N = 26, mean 46.04). Most specimens with separated prefrontal scales (93%) by an average of 0.32 mm (N = 50, 0.05 – 0.64). Ear aperture smaller (N = 62, mean 1.01, 0.50 - 1.44), than palpebral disc (N = 62, mean 1.31, 1.05 – 1.59), with many small lobules (mean 10, 5 - 16). Lamellae under third finger 9 – 20 (N = 63 mean 16.75), fourth toe 15 – 27 (N = 63 mean 22.83) (Table S3.7). Most specimens are dorsally dark brown and ventrally yellow but with either a bright or dark blue gular.
Distribution. Distributed across the sub-humid area in the Kimberley, from the northeast Berkeley River region, to the southwest King Leopold Ranges (Fig. 3.1). Present in humid islands in the Kimberley, mostly the northern islands and those closer to the mainland. In drier environments, this species tends to be more restricted to mesic microhabitats in rocky gorges (Russell Barrett pers. comm.).
Remarks. The previous described paratype from East Montalivet Island (WAM R41462) in Storr (1974) by geographic location should belong to *C. insularis* sp. nov.

Carlia insularis sp. nov. (Figs. S7A, S8A, C and S9A) urn:lsid:zoobank.org:act:F058DFD2-799C-4242-8926-9F59AEC6FD44
Kimberley islands rainbow-skink

Holotype. WAM R158646, from North Maret island, Western Australia, in -14.3983 124.97750. Specimen collected in 2004 by Richard How (Fig. S3.7A).

Paratypes. Fenelon Island: WAM R117708, WAM R117709, WAM R117710; Corneille Island: WAM R117967; West Montalivet Island: WAM R158562, WAM R158571; Don Island: WAM R158610; North Maret Island: WAM R158647 (Table S3.1, Fig. S3.8A, C).

Etymology. Insularis is derived from the Latin word insular, for island, since this species is restricted to islands.

Diagnosis. Morphologically similar to *C. johnstonei* and distinguished from this species by the presence of mid-dorsal body scales with a mix of two or three keels (Fig. 3.5), whereas *C. johnstonei* always has two keels. As mentioned previously, it is also distinguished from *C. johnstonei* by longer body size, higher relative head depth, longer relative limb length, more sharp lobules in the ear aperture (mean values of 13 vs. 10; Fig. 3.5) and more lamellae under longest finger and toe (average 3 more). Prefrontal scales are either narrowly separated or in contact, while *C. johnstonei* often has more widely separated prefrontals. From a genetic perspective, four sites that change amino acids in the mtDNA ND4 sequence reliably distinguish *Carlia johnstonei* and *Carlia insularis* sp. nov. (Table 3.2). Geographically distinct from *C. johnstonei* in some of the most outer islands of the Bonaparte Archipelago (see bellow).

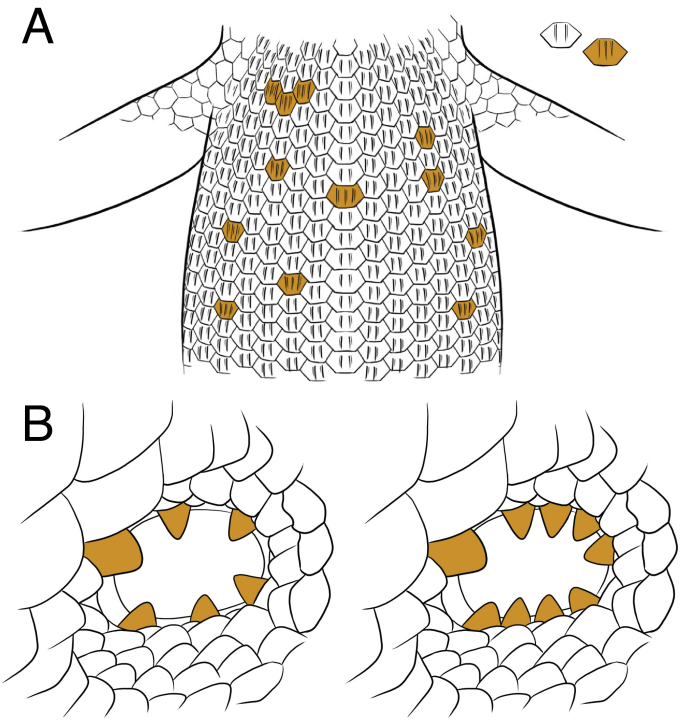


Figure 3.5 Relevant diagnostic traits. Irregular keeling in dorsal body scales for *C. insularis* sp. nov. (A) and difference in ear lobules of *C. triacantha* (left) and *C. isostriacantha* sp. nov. (right). Illustrations by Erin Walsh.

Comparison with congeners. Distinguished from remaining Australian *Carlia* species by a reduced upper preocular and well separated from posterior margin of second loreal scale (Hoskin & Couper, 2012); a distinct interparietal, usually seven supraciliaries, prefrontals usually separated; at least 34 mid-body scale rows, that are dorsally 6-sided, each scale with an angular free edge and strongly bicarinate, with the keels aligned to form continuous longitudinal lines; ear-opening surrounded by many small and pointed lobules (Cogger, 2014). It is endemic to Kimberley islands where *C. johnstonei* and *C. isostriacantha* sp. nov. also occur at a regional scale. See diagnosis to distinguish from *C. johnstonei*; and distinguishable from *C. isostriacantha* sp. nov. by the presence of two

keeled-scales and usually seven supraciliaries instead of six.

Description of holotype. Individual with 42.01 mm as SVL, tail 69.33 mm, axilla-groin length 19.71 mm, head length 8.86 mm, head width 6.29 mm, head depth 3.85 mm, forelimb 12.47 and hindlimb 17.42 mm. Body with keeled dorsal scales, mostly two keels but some scales with three. Six supraciliares, seven supralabials, six infralabials, 19 subdigital lamellae in 3rd finger, 26 subdigital lamellae in 4th toe. Circular ear not smaller (1.37 mm) than palpebral disc (1.19 mm) with 12 sharp ear lobules. Prefrontals narrowly separately and nasals widely spaced (2.56 mm). Midbody scale rows 37, 43 vertebral scales and 62 ventral scales.

Description. Snout-vent length (mm): 27.93 - 51.44 (N = 35, mean 41.83). Tail: 29.05 – 69.98 (N = 18, mean 51.02). Most specimens with separated prefrontal scales (62%) by an average of 0.18 mm (N = 21, 0.02 – 0.54). Ear aperture smaller (N = 32, mean 1.27, 0.85 – 2.16), than palpebral disc (N = 32, mean 1.44, 1.04 – 2.13), with many small lobules (up to 18). Lamelae under third finger 17 – 23 (N = 35 mean 19.69), fourth toe 21 – 30 (N = 35 mean 26.31) (Table S3.7). Laterally and dorsally blackish brown while ventrally yellowish with sometimes a bright blue or a dark blue gular (Fig. S3.8C), where in breeding males (Fig. S3.9A) lateral midbody has a light brown almost orange colour.

Distribution. Across the northwest and outer islands of the Bonaparte Archipelago (northern Kimberley islands in Western Australia) with confirmed occurrence on the Fenelon, Corneille, East Montalivet, West Montalivet, Don, Berthier, North Maret and South Maret islands.

Remarks. Despite extensive sampling, there are no records of *C. insularis* sp. nov. and *C. johnstonei* occurring on the same islands. All islands where the former species is confirmed are either laterite or volcanic islands, whereas *C. johnstonei* also occurs in sandstone islands (How et al., 2006). The individuals of *C. insularis* sp. nov. were collected in vine thicket and deciduous vine forest habitats (Richard How; pers. comm.). Despite Descartes island being relatively close to Fenelon and Corneille islands, only *C. johnstonei* was confirmed on this island.

Carlia triacantha Mitchell, 1953, Records of the South Australian Museum, Vol. 11, 75-90
Desert rainbow-skink

Holotype. SAM R2697, from Adelaide River, Northern Territory, in -13.183 131.1.

Diagnosis. Species morphologically distinguished from congeners by having three strong keels in scales, prefrontals more often in contact or very narrowly separated and usually six supraciliaries. Although more work is still needed to find unambiguously diagnostic traits between this species and *C. isostricantha* sp. nov., *C. triacantha* are mostly smaller (mean 36.55 mm vs. 40.07 mm), with shorter relative head length (mean 7.24 mm vs. 8.25 mm) and fewer ear lobules (usually 6 vs. 9, Fig. 3.5B). Geographically diagnosis from *C. isostricantha* sp. nov., possible in the centre of Australia, particularly Pilbara and Macdonald ranges region.

Description. Snout-vent length (mm): 23.78 - 44.98 (N = 35, mean 36.55). Tail: 38.48 – 75.90 (N = 17, mean 60.80). Prefrontal in contact (63%) while the rest with separated prefrontals (N = 11) by an average of 0.26 mm (0.03 – 1.81). Ear aperture smaller (N = 30, mean 1.13,

0.64 - 1.73), than palpebral disc (N = 30, mean 1.41, 0.99 - 1.71), with often one larger anterior lobule and several small (up to 7). Lamellae under third finger 16 – 22 (N = 30 mean 18.83), fourth toe 23 – 28 (N = 29 mean 24.83) (Table S3.7). Dorsally brown and ventrally yellow blueish with sometimes whitish line under eye.

Distribution. Widely distributed from Pilbara in Western Australia to Northern Territory (Fig. 3.1). However, more sampling and genetic analyses are needed to investigate whether this species is continuously distributed from the mesic Top End to arid central Australia or if the central Top End is only occupied by *C. isostriacantha* sp. nov.

Carlia isostriacantha sp. nov. (Fig. S3.7B, S3.8B,D and S3.9B) urn:lsid:zoo-bank.org:act:EB2E9D69-8E1F-466D-8441-E2E4DD59F96E
Monsoonal three-keeled rainbow-skink

Holotype. WAM R171420, from Prince Regent Nature Reserve, Western Australia, in -15.98972 125.32944. Specimen collected in 2010 by Paul Doughty (Fig. S7B).

Paratypes. WAM R168173 (Boongaree Island), WAM R168675 (Katers Island), WAM R171211 (Darcy Island), WAM R171905 (Wargul Wargul Island), WAM R171906 (Molema Island), WAM R171908 (Sunday Island), WAM R171909 (Balami ridge), WAM R171916 (Lachlan Island), WAM R171921 (Storr Island), WAM R171933 (Balami ridge) (Table S3.1, Fig. S3.8B,D).

Etymology. Isostriacantha is derived from equal in greek (isos) with triacantha, (three spines, referring to the three keels in scales) due to the difficulty of morphologically distinguishing from its sister species *C. triacantha*.

Diagnosis. As similar to *C. triacantha*, this species is morphologically distinguished from other *Carlia* species by having three strong keels in scales, prefrontals more often in contact or very narrowly separated and usually six supraciliaries. As above-mentioned, in contrast with it closest relative, *C. triacantha*, this species has longer body size, a relatively longer head and tends to have more ear lobules, on average nine very small lobules (Fig. 3.4 and 3.5, Table S3.7). Another possible trait to distinguish between these species is a white line that begins posterior to each hind limb and can extend to midway through the tail (Fig. S3.9B). This trait is more evident in freshly caught individuals, or photographs of them, than in long preserved specimens and needs to be further tested through more observations on genetically typed individuals. Genetically diagnosed from *C. triacantha*, by three ND4 mtDNA sites (Table 3.2) and geographically by occurring in the Kimberley, although geographic diagnoses in Northern Territory requires further work.

Comparison with congeners. This species can be separated from most Australian *Carlia* species by an upper preocular reduced and well separated from posterior margin of second loreal scale (Hoskin & Couper, 2012); a distinct interparietal, with usually six supraciliaries, prefrontals usually in contact or narrowly separated; 28-36 rows of mid-body scales, that are dorsally 6-sided triscupid, each usually with an angular free edge and strongly keeled; often one larger anterior lobule with many small lobules in a round ear-opening that is smaller than palpebral disc, while the palpebral disc occupies much more than half of lower eyelid

(Cogger 2014). Specifically with potentially sympatric species, *C. johnstonei*, *C. amax*, *C. rufilatus*, *C. gracilis* and *C. munda*, this species can be identified by the presence of three strong keels in scales, prefrontals usually in contact, six supraciliaries and absence of white lateral line anterior to the forelimbs. To distinguish from its sister species, *C. triacantha*, see Diagnosis above.

Description of holotype. Male individual with 43.22 mm as SVL, tail 63 mm, axilla-groin length 20.06 mm, head length 8.79 mm, head width 6.57 mm, head depth 3.83 mm, forelimb 13.06 mm and hindlimb 19.34 mm. Body with three keeled dorsal scales. Six supraciliares, seven supralabials, six infralabials, 17 subdigital lamellae in third finger, 23 subdigital lamellae in fourth toe. Horizontal ear wider (1.83 mm) than palpebral disc (1.57 mm) with 13 small sharp ear lobules (one anterior larger). Prefrontals in contact and nasals widely spaced (2.34 mm). Midbody scale rows 35, 38 vertebral scales and 52 ventral scales.

Description. Snout-vent length (mm): 24.72 - 49.12 (N = 68, mean 40.07). Tail: 29.1 – 86.68 (N = 39, mean 61.55). Prefrontal in contact (73%) while the rest with separated prefrontals (N = 19) by an average of 0.12 mm (0.01 – 0.37). Ear aperture smaller (N = 67, mean 1.33, 0.69 - 1.96), than palpebral disc (N = 67, mean 1.46, 0.92 - 1.81), with often one larger anterior lobule, many small (up to 13) and sometimes one superior. Lamelae under third finger 11 – 24 (N = 62 mean 19.27), fourth toe 18 – 30 (N = 62 mean 24.82) (Table S3.7). Dorsally brown and ventrally yellow blueish, with a light line under eye to ear, and often with a very light whitish line in the back of hindlimbs to tail if not regrown (Fig. S3.9B).

Distribution. Widespread across the Kimberley and adjacent (mostly southern Kimberley) islands in Western Australia, with isolated records in the western Gulf region, spanning the border of the Northern Territory and Queenland (Fig. 3.1).

Remarks. Afonso Silva et al. (2017) found one genetically discordant sample with mtDNA of *C. isostriacantha* sp. nov. and nuclear of *C. triacantha* from the Victoria River region (ABTC61613, Table S3.1). This suggests a need for further regional surveys and genetic studies, particularly in the Northern Territory where only a few specimens with tissues were detected, to define the boundaries of both species, at geographical and morphological level.

3.4 Discussion

We used extensive genetic and morphological data to identify two new species of Rainbow skinks, *Carlia insularis* sp. nov. (Johnstonei B lineage) and *Carlia isostriacantha* sp. nov., (Triacantha B lineage), in an understudied region of Australia, the Kimberley. We also redefined diagnoses and geographic distributions of *Carlia johnstonei* and *C. triacantha*. Our work takes advantage of recent progress in techniques for obtaining large-scale sequence data and in methods for species delimitation, as part of a broader integrative taxonomic approach. These advances are particularly important for identifying cryptic species, such as those described here, where morphological evidence alone is often insufficient for reliable species identification.

3.4.1 Evidence for new cryptic species

A previous phylogeographic study with >2000 loci Afonso Silva et al. (2017) revealed two new candidates species in the *Carlia* genus. The current work confirms these are new species using three robust hypothesis-driven validation methods based on several independent sets of genes from the larger exon dataset. The use of multiple different methods provides a robust test for the previous discovery in Afonso Silva et al. (2017), and further validates the proposed species delimitation.

Although the existence of *C. insularis* sp. nov. and *C. isostriacantha* sp. nov. is well supported in the genetic data, distinguishing these species morphologically is more difficult due to their cryptic nature. The genus *Carlia* generally has few diagnostic taxonomic characters that allow for the separation of species using morphology. Even for *C. johnstonei* and *C. triacantha* as currently recognised, there are only a few morphological characters that effectively distinguish between these species, mainly the number of keels on the dorsal scales and the arrangement of ear lobules (Storr, 1974). However, morphological measurements broadly overlap between both *C. johnstonei* and *C. triacantha* lineages. Despite these issues, we were able to find statistically significant differences in morphology across both measurements and meristic data, supporting the presence of these lineages as different species.

Differences in body size, head and limbs traits as well as ear lobule numbers help in distinguishing the lineages. Morphological variation across each pair of taxa is strongly affected by body size (SVL), with the newly described species being larger than their respective sister taxa. The same is observed for the other significant traits, even after accounting for size. Although for both species, there are some overlap between morphological groups, there was more morphological similarity between the *C. triacantha* lineages than between *C. johnstonei* lineages (Fig. 3.3), likely reflecting the shallower divergence seen within *C. triacantha*.

Though we were able to identify a few distinct morphological traits, using morphology alone to identify individuals will remain a challenge without a reference to geography. For the two lineages within *C. triacantha*, even geography is a poor guide for the central Northern Territory region. Therefore, for more reliable diagnosis, we follow the suggestion of Renner (2016) and include a set of diagnostic mtDNA SNPs to distinguish between *C. johnstonei* and *C. insularis* sp. nov., and between *C. triacantha* and *C. isostriacantha* sp. nov. These SNPs can be easily assessed by cheaper Sanger sequencing of the mtDNA gene ND4 (primers and protocol in Afonso Silva et al., 2017).

3.4.2 Biodiversity significance of the two new species

C. insularis sp. nov. is an important addition to the known biodiversity of the Kimberley islands. This region has recently been the focus of several studies that have documented

unique biodiversity communities, namely in terms of vegetation (Lyons et al., 2014), avifauna (Pearson & Caton, 2013) and herpetofauna (Doughty et al., 2012; Palmer et al., 2013). Studies to understand the biodiversity value in this region are also of importance to conservation, as this area is being considered as a biodiversity refuge for fauna vulnerable to the invasive Cane Toad (Palmer et al., 2013). Although the west Kimberley region has several endemic species, only a few are endemic just to the islands, namely a blindsnake (Ellis, 2016) and several land snails (Criscione & Köhler, 2013, 2014), making the discovery of *C. insularis* sp. nov. very significant. But more island-endemics reptiles are expected to be described, since Palmer et al. (2013) suggested the occurrence of a few potential new species (including samples that correspond to *C. insularis* sp. nov.) that have not yet been described.

Although our genetic data allows us to describe *C. isostriacantha* sp. nov. as a new species, further collecting and analyses are needed across central Northern Territory for this and other taxa (also suggested in Rosauer et al., 2016). Specifically, there is a need to identify the geographic distributions of *C. triacantha (s. s.)* and *C. isostriacantha* sp. nov., as well as to examine morphological divergence in this poorly sampled region. In a group of Ctenotus skinks, Rabosky et al. (2014) highlight how intraspecific morphological variability and geographic sampling gaps caused an inadequate understanding of biological diversity. As with Ctenotus, we suspect that many other species in the *Carlia* genus may yet require taxonomic revision. Potter et al. (2016) have also suggested unknown lineage diversity in another *Carlia* species in the Australian Monsoonal Tropics, which may lead to the description of additional *Carlia* species, particularly on the islands off the northeast Top End.

3.4.3 Advantages and issues of using MSC methods

A key element of our analysis was the use of multispecies coalescent (MSC) methods, including pioneering the application of StarBEAST2 to Bayes Factor species delimitation (BFD). MSC models are a robust approach that better describes species formation by considering coalescent processes; however, methods based on the MSC are typically computationally intensive. To surpass this limitation, we subsetted independent smaller sets of loci from around 2,300 loci, which also has the advantage of producing multiple replicate results that may be compared to confirm that estimated parameter values are robust to the choice of loci.

BFD using SNAPP and StarBEAST2 requires sampling from different power posteriors, including sampling purely from the prior. We found that convergence was difficult to achieve for our data set when BFD StarBEAST2 was used to sample from the prior with more than 20 loci. Despite this limitation, BFD StarBEAST2 has advantages over existing methods for species delimitation. Compared to SNAPP which requires unlinked SNPs, StarBEAST2 can extract much more information from each locus. Compared to BPP, StarBEAST2 has many more options for substitution models, population size models, and relaxed clock models.

3.5 Conclusions

As Oliver, Keogh & Moritz (2015) express, most genetically divergent lineages within species remain invisible to other scientific work, like conservation assessments and management planning. This reinforces the need to evaluate whether genetically distinct lineages within species should be formally described. Here we validate and describe two new species of rainbow skinks in the northwest of Australia, a highly biodiverse region of Australia that is still relatively understudied. Using an integrative taxonomic approach, we employ three MSC methods, including the application of a new approach to delimit species, as well as integrating morphological data to provide strong evidence for these two new species. This work brings the number of Australian *Carlia* to 26 species. However, further such work is needed across the Australian Monsoonal Tropics, since deeply divergent lineages within species of lizards are the norm in this region.

3.6 References

Abràmoff MD., Magalhães PJ., Ram SJ. 2004. Image processing with ImageJ. Biophotonics international 11:36–42.

Afonso Silva AC., Bragg JG., Potter S., Fernandes C., Coelho MM., Moritz C. 2017. Tropical specialist vs. climate generalist: Diversification and demographic history of sister species of *Carlia* skinks from northwestern Australia. Molecular Ecology 0:1–14. DOI: 10.1111/mec.14185.

Andersen AN., Bocciarelli D., Fairman R., Radford IJ. 2014. Conservation status of ants in an iconic region of monsoonal Australia: levels of endemism and responses to fire in the eastern Kimberley. Journal of Insect Conservation 18:137–146. DOI: 10.1007/s10841-014-9624-x.

Bickford D., Lohman DJ., Sodhi NS., Ng PKL., Meier R., Winker K., Ingram KK., Das I. 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22:148–155. DOI: 10.1016/j.tree.2006.11.004.

Blom MPK. 2015. EAPhy: A flexible tool for high-throughput quality filtering of exon-alignments and data processing for phylogenetic methods. PLOS Currents Tree of Life 1. DOI: doi:10.1371/currents.tol. 75134257bd389c04bc1d26d42aa9089f.

Bragg JG., Potter S., Bi K., Moritz C. 2015. Exon capture phylogenomics: efficacy across scales of divergence. Molecular Ecology Resources 16:1059–1068. DOI: 10.1111/1755-0998.12449.

Bryant D., Bouckaert R., Felsenstein J., Rosenberg NA., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29:1917–1932. DOI: 10.1093/molbev/mss086.

Carstens BC., Pelletier TA., Reid NM., Satler JD. 2013. How to fail at species delimitation. Molecular Ecology 22:4369–4383. DOI: 10.1111/mec.12413.

Cogger H. 2014. Reptiles and amphibians of Australia. New Holland, Sydney.

Criscione F., Köhler F. 2013. Six new species of Australocosmica Köhler, 2011 from the Kimberley Islands, Western Australia (Mollusca: Pulmonata: Camaenidae). Zootaxa 3608:101–15. DOI: 10.11646/zootaxa.3608.2.1.

Criscione F., Köhler F. 2014. Molecular phylogeny and taxonomic revision of the genera Baudinella Thiele, 1931, Retroterra Solem, 1985 and Molema Köhler, 2011 endemic to the coastal Kimberley, Western Australia (Gastropoda, Camaenidae). Journal of Zoological Systematics and Evolutionary Research 52:273–284. DOI: 10.1111/jzs.12065.

Darriba D., Taboada GL., Doallo R., Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:772. DOI: 10.1038/nmeth.2109.

Degnan JH., DeGiorgio M., Bryant D., Rosenberg NA. 2009. Properties of Consensus Methods for Inferring Species Trees from Gene Trees. Systematic Biology 58:35–54. DOI: 10.1093/sysbio/syp008.

Dolman G., Hugall AF. 2008. Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). Molecular Phylogenetics and Evolution 49:782–94. DOI: 10.1016/j.ympev.2008.09.021.

Doughty P., Palmer R., Cowan M., Pearson DJ. 2012. Biogeographic patterns of frogs of the Kimberley islands, Western Australia. Records of the Western Australian Museum Supplement 81:109–124. DOI: 10.18195/issn.0313-122x.81.2012.109-124.

Drummond AJ., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214. DOI: 10.1186/1471-2148-7-214.

Edwards S V., Xi Z., Janke A., Faircloth BC., McCormack JE., Glenn TC., Zhong B., Wu S., Lemmon EM., Lemmon AR., Leaché AD., Liu L., Davis CC. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Molecular Phylogenetics and Evolution 94:447–462. DOI: 10.1016/j.ympev.2015.10.027.

Ellis RJ. 2016. A New Species of Blindsnake (Scolecophidia: Typhlopidae: Anilios) from the Kimberley Region of Western Australia. Herpetologica 72:271–278. DOI: 10.1655/Herpetologica-D-16-00007.1.

Ence DD., Carstens BC. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. Molecular Ecology Resources 11:473–80. DOI: 10.1111/j.1755-0998.2010.02947.x.

Fujita MK., Leaché AD., Burbrink FT., McGuire JA., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in ecology & evolution 27:480–8. DOI: 10.1016/j.tree.2012.04.012.

Grummer JA., Bryson RW., Reeder TW. 2014. Species delimitation using Bayes Factors: simulations and application to the Sceloporus scalaris species group (Squamata: Phrynosomatidae). Systematic Biology 63:119–133. DOI: 10.1093/sysbio/syt069.

Guindon S., Gascuel O., Rannala B. 2003. A Simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696. DOI: 10.1080/10635150390235520.

Hedin M., Carlson D., Coyle F. 2015. Sky island diversification meets the multispecies coalescent - divergence in the spruce-fir moss spider (Microhexura montivaga, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. Molecular Ecology 24:3467–3484. DOI: 10.1111/mec.13248.

Hoskin CJ. 2014. A new skink (Scincidae: *Carlia*) from the rainforest uplands of Cape Melville, north-east Australia. Zootaxa 3869:224–236. DOI: 10.11646/zootaxa.3869.3.2.

Hoskin CJ., Couper PJ. 2012. Description of two new *Carlia* species (Reptilia: Scincidae) from north-east Australia, elevation of *Carlia* pectoralis inconnexa Ingram & Covacevich 1989 to full species status, and redescription of *Carlia* pectoralis (de Vis 1884). Zootaxa:1–28.

How R., Schmitt L., Teale R., Cowan M. 2006. Appraising vertebrate diversity on Bonaparte Islands, Kimberley, Western Australia. The Western Australian Naturalist 25:92–110.

Isaac NJB., Mallet J., Mace GM. 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology & Evolution 19:464–469. DOI: 10.1016/j.tree.2004.06.004.

Kass RE., Raftery AE. 1995. Bayes Factors. Journal of the American Statistical Association 90:773–795. DOI: 10.1080/01621459.1995.10476572.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–9. DOI: 10.1093/bioinformatics/bts199.

Köhler F. 2011. Descriptions of new species of the diverse and endemic land snail Amplirhagada iredale, 1933 from rainforest patches across the Kimberley, Western Australia (Pulmonata:

Camaenidae). Records of the Australian Museum 63:167–202. DOI: 10.3853/j.0067-1975.63.2011.1581.

Komsta L. 2011. Outliers: Tests for outliers. R package version 0.14. https://CRAN.R-project.org/package=outliers.

Korkmaz S., Goksuluk D., Zararsiz G. 2014. MVN: An R package for assessing multivariate normality. The R Journal 6:151–162.

Lanier HC., Knowles LL. 2012. Is Recombination a problem for species-tree analyses? Systematic Biology 61:691. DOI: 10.1093/sysbio/syr128.

Leaché AD., Fujita MK. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). Proceedings of the Royal Society of London B: Biological Sciences 277:3071–3077. DOI: 10.1098/rspb.2010.0662.

Leaché AD., Fujita MK., Minin VN., Bouckaert RR. 2014. Species delimitation using genome-wide SNP data. Systematic Biology 63:534–42. DOI: 10.1093/sysbio/syu018.

Liaw A., Wiener M. 2002. Classification and regression by randomForest. R News 2:18–22.

Lyons MN., Keighery GJ., Gibson LA., Handasyde T. 2012. Flora and vegetation communities of selected islands off the Kimberley coast of Western Australia. Records of the Western Australian Museum Supplement 81:205–244. DOI: 10.18195/issn.0313-122x.81.2014.205-244.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–303. DOI: 10.1101/gr.107524.110.

Mitchell FJ. 1953. A brief revision of the four-fingered members of the genus Leiolopisma. Records of the South Australian Museum:75–90.

Ogilvie HA., Bouckaert R., Drummond AJ. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Molecular Biology and Evolution. DOI: 10.1093/molbev/msx126.

Oliver PM., Bourke G., Pratt RC., Doughty P., Moritz C. 2016. Systematics of small Gehyra (Squamata: Gekkonidae) of the southern Kimberley, Western Australia: Redescription of G. kimberleyi Börner & Schüttler, 1983 and description of a new restricted range species. Zootaxa 4107:49–64. DOI: 10.11646/zootaxa.4107.1.2.

Oliver P., Keogh JS., Moritz C. 2015. New approaches to cataloguing and understanding evolutionary diversity: a perspective from Australian herpetology. Australian Journal of Zoology 62:417–430. DOI: 10.1071/ZO14091.

Oliver PM., Laver RJ., Melville J., Doughty P. 2014. A new species of Velvet Gecko (Oedura: Diplodactylidae) from the limestone ranges of the southern Kimberley, Western Australia. Zootaxa 3873:49–61. DOI: 10.11646/zootaxa.3873.1.4

Padial J., Miralles A., Riva ID la., Vences M. 2010. The integrative future of taxonomy. Frontiers in Zoology 7:1–14. DOI: 10.1186/1742-9994-7-16

Palmer R., Pearson DJ., Cowan MA., Doughty P. 2013. Islands and scales: a biogeographic survey of reptiles on Kimberley islands, Western Australia. Records of the Western Australian Museum Supplement 81:183–204. DOI: 10.18195/issn.0313-122x.81.2013.183-204

Park J., Koo K-S., Kim I-H., Park D. 2016. Complete mitochondrial genomes of Scincella vandenburghi and S. huanrenensis (Squamata: Scincidae). Mitochondrial DNA Part B 1:237–238. DOI: 10.1080/23802359.2016.1156490.

Pearson DJ., Caton W. 2013. The avifauna of larger islands along the Kimberley coast, Western Australia. Records of the Western Australian Museum Supplement 81:125–144. DOI: 10.18195/issn.0313-122x.81.2013.125-144.

Phillips BL., Baird SJE., Moritz C. 2004. When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia* rubrigularis. Evolution 58:1536–1548. DOI: 10.1111/j.0014-3820.2004.tb01734.x.

Potter S., Bragg JG., Peter BM., Bi K., Moritz C. 2016. Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia* amax. Molecular Ecology 25:1367–1380. DOI: 10.1111/mec.13546.

Rabosky DL., Hutchinson MN., Donnellan SC., Talaba AL., Lovette IJ. 2014. Phylogenetic disassembly of species boundaries in a widespread group of Australian skinks (Scincidae: Ctenotus). Molecular Phylogenetics and Evolution 77:71–82. DOI: 10.1016/j.ympev.2014.03.026.

R Core Team. 2016. A language and environment for statistical computing.

Rambaut A., Suchard MA., Xie D., Drummond AJ. 2015. Tracer v1.6., Available from http://beast.bio.ed.ac.uk/Tracer.

Rannala B. 2015. The art and science of species delimitation. Current Zoology 61:846–853. DOI: 10.1093/czoolo/61.5.846

Rannala B., Yang Z. 2013. Improved reversible jump algorithms for bayesian species delimitation. Genetics 194:245–253. DOI: 10.1534/genetics.112.149039.

Renner SS. 2016. A return to Linnaeus's focus on diagnosis, not description: the use of DNA characters in the formal naming of species. Systematic Biology 65:1085–1095. DOI: 10.1093/sysbio/syw032.

Ripley B., Venables B., Bates DM., Hornik K., Gebhardt A., Firth D., Ripley MB. 2013. Package "MASS." Cran R.

Rosauer DF., Blom MPK., Bourke G., Catalano S., Donnellan S., Gillespie G., Mulder E., Oliver PM., Potter S., Pratt RC., Rabosky DL., Skipwith PL., Moritz C. 2016. Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia. Biological Conservation 204:83–93. DOI: 10.1016/j.biocon.2016.05.002.

Rosenblum EB., Sarver B a J., Brown JW., Des Roches S., Hardwick KM., Hether TD., Eastman JM., Pennell MW., Harmon LJ. 2012. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. Evolutionary Biology 39:255–261. DOI: 10.1007/s11692-012-9171-x.

Singhal S., Moritz C. 2013. Reproductive isolation between phylogeographic lineages scales with divergence. Proceedings of the Royal Society B: Biological Sciences 280:20132246. DOI: 10.1098/rspb.2013.2246.

Storr GM. 1974. The genus *Carlia* (Lacertilia, Scincidae) in Western Australia and Northern Territory. Records of the Western Australian Museum 3:151–165.

Sukumaran J., Knowles LL. 2017. Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences 2016:201607921. DOI: 10.1073/PNAS.1607921114.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer.

Woerner AE., Cox MP., Hammer MF. 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics 23:1851–1853. DOI: 10.1093/bioinformatics/btm253.

Xie W., Lewis PO., Fan Y., Kuo L., Chen M-H. 2010. Improving marginal likelihood estimation for bayesian phylogenetic model selection. Systematic Biology 60:150. DOI: 10.1093/sysbio/syq085.

Yang Z. 2015. The BPP program for species tree estimation and species delimitation. Current Zoology 61:854. DOI: 10.1093/czoolo/61.5.854.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences of the United States of America 107:9264–9. DOI: 10.1073/pnas.0913022107.

Yang Z., Rannala B. 2014. Unguided species delimitation using DNA sequence data from multiple loci. Molecular Biology and Evolution 31:3125–3135. DOI: 10.1093/molbev/msu279.

Acknowledgements

We thank the Museum and Art Gallery of the Northern Territory, Museum Victoria, South

Australia Museum and Western Australia Museum for specimens, in particular to Ryan Ellis for his assistance. We are also grateful to members of the Moritz lab for ongoing discussions and comments relating to this study, and in particular to Mozes Blom, Joshua Peñalba, Rebecca Laver, Sally Potter, as well to Emma Sherratt, Liam Bailey and Thomas Merkling for advice on analysis and comments on the manuscript. We are also grateful to Fabricius Domingos and an anonymous reviewer for helpful comments. We acknowledge credit and thank to Erin Walsh, Damien Esquerré, Russell Barrett and Mark Allen for illustrations and photos.

Additional information and declarations

Funding

This research was supported by grants from the Australian Biological Resources Study to CM and Scott Keogh, and from the Australian Research Council to CM (ARC FL110100104). ACAS is supported by the FCT grant SFRH/BD/88740/2012. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Ana C. Afonso Silva conceived and designed the experiments, performed the experiments, analysed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper. Natali Santos performed the experiments, analysed the data. Huw A. Ogilvie conceived and designed the experiments, performed the experiments, analysed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper. Craig Moritz conceived and designed the experiments, contributed reagents/materials/- analysis tools, wrote the paper, reviewed drafts of the paper.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers): The Australian National University Animal ethics office provided full approval for this research permit with the number A2012/14.

Data Availability

The following information was supplied regarding data availability: Silva, Ana (2017): Da-ta and code to describe *Carlia johnstonei* and *C. triacantha*. figshare. https://doi.org/10.6084/m9.figshare.4621963.v3.

New Species Registration

The following information was supplied regarding the registration of a newly described species: Publication LSID: urn:lsid:zoobank.org:pub:A7B29F16-079F-48BA-B4BE-3EC9A3D80D34. *Carlia insularis*: LSID: urn:lsid:zoobank.org:act:F058DFD2-799C-4242-8926- 9F59AEC6FD44. *Carlia isostriacantha*: LSID: urn:lsid:zoobank.org:act:EB2E9D69-8E1F-466D-8441- E2E4DD59F96E.

3.7 Supplementary material

Table S3.1 Tissue and specimens list with museum origin, mtDNA lineage, for which analysis samples were used (SD – species delimitation, M – morphology), sex information (F – female, M – male, NA – not available) if available and location. * Correspond to the genetically discordant sample with mtDNA of Triacantha B but nuclear of Triacantha A (with no evidence of admixture in Afonso Silva et al., 2017). WAM - Western Australian Museum; ANU – Australian National University; MV – Museum Victoria; SAM – South Australian Museum; NTM – Northern Territory Museum. JA – Johnstonei A; JB – Johnstonei B; TA – Triacantha A; TB – Triacantha B.

Tissue/ Specimen	Specimen location	mtDNA lineage	Analysis	Sex	Latitude	Longitude	Location
BP00664	WAM	JA	SD	NA	-14.97630	124.91310	Coronation Island
BP01423	WAM	JA	SD	NA	-15.94970	124.56110	Storr Island
CCM0737	ANU	JA	M	NA	-15.19890	125.90350	Spring Creek
CCM0762	ANU	JA	M	NA	-15.35220	126.58810	Carson escarpment
CCM0763	ANU	JA	M	NA	-15.33810	126.58910	Carson escarpment
CCM0933	ANU	JA	M	F	-14.51903	126.45811	Theda Station
CCM0999	ANU	JA	M	F	-14.88556	126.35986	Theda Station
CCM1125	ANU	JA	SD/M	F	-14.78060	126.63690	Langoor Hill
CCM1126	ANU	JA	M	F	-14.78060	126.63690	Theda Station
CCM1127	ANU	JA	M	F	-14.78680	126.63420	Theda Station
CCM1128	ANU	JA	M	F	-14.78680	126.63420	Theda Station
CCM1132	ANU	JA	M	M	-14.78290	126.63510	Theda Station
CCM1133	ANU	JA	M	M	-14.78290	126.63510	Langoor Hill
CCM1195	ANU	JA	M	F	-14.82110	125.72140	Mitchell Plateau
CCM1196	ANU	JA	M	F	-14.82110	125.72140	Mitchell Plateau
CCM1197	ANU	JA	M	NA	-14.82110	125.72110	Mitchell Plateau
CCM1200	ANU	JA	M	F	-14.83170	125.71920	Mitchell Plateau
CCM1264	ANU	JA	M	F	-17.06837	125.24618	Silent Grove
CCM1290	ANU	JA	M	F	-16.78641	124.92006	Mt Matthew Gorge
CCM1301	ANU	JA	M	F	-16.97111	125.02998	Mt Hart
CCM1499	ANU	JA	M	NA	-16.99694	125.20469	Bell Gorge Creek
CCM1761	ANU	JA	M	F	-14.79721	126.50038	Theda Station
CCM1762	ANU	JA	M	F	-14.79721	126.50038	Theda Station
D76985	MV	JA	M	NA	-16.99920	125.20680	Bell Gorge
D76986	MV	JA	M	NA	-16.99920	125.20680	Bell Gorge
R106285	WAM	JA	M	NA	-15.43333	124.60000	Augustus Island
R113992	WAM	JA	M	NA	-15.35000	126.61667	Carson Escarpment
R117719	WAM	JA	M	M	-14.16389	125.67972	Descartes Island
R117726	WAM	JA	M	NA	-14.16380	125.67970	Descartes Island
R117760	WAM	JA	M	NA	-14.16389	125.67972	Descartes Island
R117772	WAM	JA	M	NA	-14.38020	125.95110	South West Osborn Island
R117787	WAM	JA	M	NA	-14.38028	125.95111	South West Osborn Island
R117793	WAM	JA	M	M	-14.37972	125.94917	South West Osborn Island
R117799	WAM	JA	M	NA	-14.34333	126.02083	Middle Osborn Island
R117801	WAM	JA	M	NA	-14.34194	126.02111	Middle Osborn Island
R117811	WAM	JA	M	NA	-14.37556	125.98972	*Carlia* Island
R117812	WAM	JA	M	NA	-14.33833	125.97694	Kidney Island
R117813	WAM	JA	M	NA	-14.33833	125.97694	Kidney Island
R117822	WAM	JA	M	NA	-14.37667	125.98972	*Carlia* Island
R117838	WAM	JA	M	NA	-14.37550	125.98970	*Carlia* Island
R117843	WAM	JA	M	NA	-14.34917	125.95917	South West Osborn Island
R117844	WAM	JA	M	NA	-14.34910	125.95920	South West Osborn Island
R117851	WAM	JA	M	M	-14.34917	125.95917	South West Osborn Island
R117855	WAM	JA	M	NA	-14.44694	125.99722	Steep Head Island
R158783	WAM	JA	M	F	-14.60194	125.20389	Bigge Island
R158784	WAM	JA	M	F	-14.60194	125.20389	Bigge Island
R158915	WAM	JA	M	F	-14.61833	125.24111	Purrungku Island
R167192	WAM	JA	M	F	-15.37139	124.94333	Coronation Island
R167235	WAM	JA	M	NA	-15.04889	124.95000	Gray Island
R167242	WAM	JA	M	M	-15.06278	124.96194	Glauert Island

Tissue/ Specimen	Specimen location	mtDNA lineage	Analysis	Sex	Latitude	Longitude	Location
R167260	WAM	JA	M	F	-15.11639	124.92417	Prince Regent River
R167423	WAM	JA	M	F	-15.09222	125.12139	Boongaree Island
R167431	WAM	JA	M	M	-15.08806	125.17083	Boongaree Island
R167860	WAM	JA	M	NA	-14.82389	125.70889	Mitchell Plateau
R168219	WAM	JA	M	M	-14.60000	125.11667	Bigge Island
R168651	WAM	JA	M	NA	-14.46667	125.53333	Katers Island
R168803	WAM	JA	M	NA	-14.31667	126.00000	Middle Osborn Island
R168923	WAM	JA	M	M	-15.08333	125.20000	Boongaree Island
R171227	WAM	JA	M	M	-15.35139	124.52694	Augustus Island
R171230	WAM	JA	M	NA	-15.29056	124.39917	Darcy Island
R171237	WAM	JA	SD/M	NA	-15.25972	124.44528	Darcy Island
R171238	WAM	JA	M	M	-15.38556	124.36194	Byam Martin Island
R171242	WAM	JA	M	F	-15.38556	124.36194	Byam Martin Island
R171243	WAM	JA	M	NA	-15.35806	124.99611	Saint Andrew Island
R171488	WAM	JA	M	F	-15.99167	125.32833	Prince Regent Nature Reserve
R171897	WAM	JA	M	F	-13.93972	126.17361	Wargul Wargul Island
R171900	WAM	JA	M	M	-15.90944	124.46250	Balami ridge
R172784	WAM	JA	M	F	-16.15889	125.30722	King Leopold Ranges Cons. Park
R173481	WAM	JA	SD	NA	-14.60824	126.93172	Drysdale River
R117703	WAM	JB	M	NA	-14.18750	125.73390	Corneille Island
R117708	WAM	JB	M	NA	-14.13910	125.69940	Fenelon Island
R117709	WAM	JB	M	NA	-14.13910	125.69940	Fenelon Island
R117710	WAM	JB	M	NA	-14.13860	125.69780	Fenelon Island
R117751	WAM	JB	M	NA	-14.18880	125.73220	Corneille Island
R117945	WAM	JB	M	M	-14.18556	125.72444	Corneille Island
R117953	WAM	JB	SD/M	NA	-14.18580	125.73470	Corneille Island
R117964	WAM	JB	M	NA	-13.95160	125.64330	Cassini Island
R117967	WAM	JB	SD/M	NA	-14.18550	125.72440	Corneille Island
R158508	WAM	JB	M	NA	-14.28250	125.30470	Don Island
R158540	WAM	JB	M	F	-14.29472	125.22556	West Montalivet Island
R158543	WAM	JB	M	F	-14.29472	125.22556	West Montalivet Island
R158546	WAM	JB	M	F	-14.29470	125.22560	West Montalivet Island
R158561	WAM	JB	M	M	-14.29470	125.22560	West Montalivet Island
R158562	WAM	JB	M	M	-14.29470	125.22560	West Montalivet Island
R158564	WAM	JB	M	M	-14.29470	125.22560	West Montalivet Island
R158565	WAM	JB	M	F	-14.29470	125.22560	West Montalivet Island
R158569	WAM	JB	M	F	-14.29470	125.22560	West Montalivet Island
R158571	WAM	JB	M	F	-14.29470	125.22560	West Montalivet Island
R158572	WAM	JB	M	F	-14.29470	125.22560	West Montalivet Island
R158573	WAM	JB	M	F	-14.29470	125.22560	West Montalivet Island
R158580	WAM	JB	M	M	-14.27940	125.30610	East Montalivet Island
R158581	WAM	JB	M	F	-14.27940	125.30610	East Montalivet Island
R158582	WAM	JB	M	F	-14.27940	125.30610	East Montalivet Island
R158583	WAM	JB	SD/M	M	-14.27944	125.30611	East Montalivet Island
R158610	WAM	JB	M	F	-14.28250	125.30470	Don Island
R158646	WAM	JB	M	F	-14.39830	124.97750	North Maret Island
R158647	WAM	JB	M	F	-14.39830	124.97750	North Maret Island
R158648	WAM	JB	M	F	-14.39830	124.97750	North Maret Island
R158655	WAM	JB	M	F	-14.51660	124.98330	Berthier Island
R158657	WAM	JB	SD/M	NA	-14.51667	124.98333	Berthier Island
R158661	WAM	JB	M	F	-14.51660	124.98330	Berthier Island
R158683	WAM	JB	M	NA	-14.39830	124.97750	North Maret Island
R158705	WAM	JB	SD/M	NA	-14.39833	124.97750	North Maret Island
R165970	WAM	JB	M	F	-14.40139	124.97917	Maret Island
ABTC11723/ R34050	SAM	TA	M	NA	-23.98333	120.01667	Weelarrana
ABTC11847/ R34151	SAM	TA	M	NA	-12.65000	132.88333	Jabiru
ABTC29091/ R20877	NTM	TA	SD	NA	-12.66667	132.88333	Jabiru

Tissue/ Specimen	Specimen location	mtDNA lineage	Analysis	Sex	Latitude	Longitude	Location
ABTC29692/ R22162	SAM	TA	M	NA	-13.26528	130.96222	Litchfield
ABTC33954/ R42082	SAM	TA	M	NA	-26.34972	130.17167	Illintjitja
ABTC41790/ R46088	SAM	TA	SD/M	NA	-25.98028	129.46306	Waltitjara
CCM1859	ANU	TA	SD/M	M	-23.58338	134.49942	Ross River
CCM1860	ANU	TA	M	F	-23.58338	134.49942	Ross River Dump
CCM1888	ANU	TA	M	NA	-23.72534	133.34322	Birthday Water Hole
R106160	WAM	TA	M	M	-23.86667	118.56667	Turee Creek
R108988	WAM	TA	M	NA	-20.26944	118.89722	Lesley Salt Works
R108990	WAM	TA	M	NA	-20.28050	118.88060	Lesley Salt Works
R110273	WAM	TA	M	NA	-22.82000	119.61250	Mile Camp
R110905	WAM	TA	M	M	-19.63222	128.87528	Tanami Downs
R111810	WAM	TA	M	F	-23.49861	120.29111	Wheelarra Hill
R113038	WAM	TA	M	NA	-20.26944	118.89722	Lesley Salt Works
R129921	WAM	TA	M	NA	-23.25000	118.66667	West Angelas
R139010	WAM	TA	SD/M	NA	-19.80833	121.46389	Mandora
R139404	WAM	TA	M	NA	-21.21639	120.34056	Meentheena
R142456	WAM	TA	M	M	-22.85056	127.83250	Kiwirrkurra
R142973	WAM	TA	M	F	-22.85050	127.83250	Kiwirrkurra
R145500	WAM	TA	M	NA	-22.39000	119.99000	Port Hedland
R145556	WAM	TA	M	NA	-20.91000	118.68000	Port Hedland
R145752	WAM	TA	M	NA	-22.09000	118.99194	Chichester Range
R154566	WAM	TA	M	NA	-23.37917	120.10528	Newman
R157406	WAM	TA	M	F	-19.58917	128.86028	Tanami Desert
R157466	WAM	TA	M	M	-19.89970	128.82700	Tanami Desert
R158101	WAM	TA	M	NA	-21.60417	117.07750	Millstream
R160269	WAM	TA	M	NA	-22.46639	119.02250	Mount Marsh
R161603	WAM	TA	M	NA	-19.99833	119.35861	Goldsworthy
R164546	WAM	TA	M	NA	-21.62194	116.38972	Pannawonica
R170651	WAM	TA	M	NA	-21.62194	116.38972	Pannawonica
ABTC28797/ R16430	NTM	TB	SD/M	NA	-18.60000	136.10000	Brunette Downs
ABTC29531/ R21401	NTM	TB	SD/M	NA	-18.60833	137.98833	Lawn Hill
ABTC29554/ R21444	SAM	TB	M	NA	-18.59583	138.13183	Lawn Hill
ABTC29573/ R21478	SAM	TB	M	NA	-18.59583	138.13183	Lawn Hill
ABTC29851/ R22383	NTM	TB	SD	NA	-15.97083	129.04028	Keep River
CCM0738	ANU	TB	M	NA	-15.19890	125.90350	Spring Creek
CCM0888	ANU	TB	M	F	-14.83470	126.30013	Theda Station
CCM0932	ANU	TB	M	F	-14.73298	126.46345	Theda Station
CCM1130	ANU	TB	SD/M	F	-14.78110	126.63490	Langoor
CCM1136	ANU	TB	M	F	-14.76990	126.57880	Theda Station
CCM1199	ANU	TB	M	F	-14.83170	125.71920	Mitchell Plateau
CCM1202	ANU	TB	M	NA	-14.16710	125.73320	Mitchell Plateau
CCM1235	ANU	TB	M	NA	-17.04067	125.22680	Bell Gorge
CCM1394	ANU	TB	M	M	-16.49954	125.33637	Grevillea gorge
CCM1395	ANU	TB	M	M	-16.49954	125.33637	Grevillea gorge
CCM1643	ANU	TB	M	F	-16.17108	125.98763	Munja track
CCM1767	ANU	TB	M	F	-17.60560	126.04110	Mornington Gorge
CCM1781	ANU	TB	M	NA	-16.04730	126.70170	Russ Creek
CCM1782	ANU	TB	M	M	-17.31000	126.07700	Glenroy Meatworks
CCM1784	ANU	TB	M	NA	-16.80000	125.84430	Galvan Gorge
R108782	WAM	TB	M	NA	-17.25000	128.30000	Bream Gorge
R114408	WAM	TB	M	NA	-16.08333	123.41667	King Hall Island
R117796	WAM	TB	M	NA	-14.34194	126.02111	Middle Osborn Island

Tissue/ Specimen	Specimen location	mtDNA lineage	Analysis	Sex	Latitude	Longitude	Location
R117798	WAM	TB	M	NA	-14.34330	126.02080	Middle Osborn Island
R117853	WAM	TB	M	NA	-14.34917	125.95917	South West Osborn Island
R117946	WAM	TB	M	NA	-14.13778	125.69861	Fenelon Island
R146022	WAM	TB	M	NA	-16.68330	123.83330	Kimbolton
R151868	WAM	TB	M	NA	-13.88333	126.56667	Sir Graham Moore Island
R151958	WAM	TB	M	NA	-14.35000	125.95000	South West Osborn Island
R158008	WAM	TB	M	F	-16.14861	123.77944	Koolan Island
R158016	WAM	TB	M	F	-16.14580	123.74920	Koolan Island
R158551	WAM	TB	M	F	-14.29472	125.22556	West Montalivet Island
R158609	WAM	TB	M	F	-14.28250	125.30472	Don Island
R158970	WAM	TB	M	F	-16.14861	123.77944	Koolan Island
R164310	WAM	TB	M	F	-14.78333	125.03333	Lamarck Island
R165402	WAM	TB	M	NA	-16.03444	123.53750	Bathurst Island
R165555	WAM	TB	M	M	-16.12220	123.73440	Koolan Island
R166929	WAM	TB	M	F	-14.82972	125.71889	Mitchell Plateau
R168173	WAM	TB	M	M	-15.07270	125.17970	Boongaree Island
R168237	WAM	TB	M	NA	-16.07472	123.55000	Irvine Island
R168392	WAM	TB	M	F	-16.10500	123.51222	Gibbings Island
R168394	WAM	TB	M	F	-16.10500	123.51220	Gibbings Island
R168407	WAM	TB	M	M	-16.08360	123.54190	Irvine Island
R168426	WAM	TB	M	F	-16.14889	123.58833	Margaret Island
R168462	WAM	TB	M	NA	-13.88333	126.56667	Sir Graham Moore Island
R168590	WAM	TB	SD	NA	-14.98333	124.91667	Coronation Island
R168675	WAM	TB	M	NA	-14.46667	125.53333	Katers Island
R168707	WAM	TB	M	M	-14.46660	125.53330	Katers Island
R168762	WAM	TB	M	F	-14.31667	126.00000	Middle Osborn Island
R168910	WAM	TB	M	F	-14.60000	125.11670	Bigge Island
R169985	WAM	TB	M	F	-16.19889	123.64583	Yampi Peninsula
R169990	WAM	TB	M	F	-16.17660	123.63940	Yampi Peninsula
R171206	WAM	TB	M	M	-15.35139	124.52694	Augustus Island
R171211	WAM	TB	M	M	-15.25972	124.44528	Darcy Island
R171420	WAM	TB	M	M	-15.98972	125.32944	Prince Regent Nature Reserve
R171904	WAM	TB	M	NA	-16.62250	123.47111	Lachlan Island
R171905	WAM	TB	M	F	-13.93972	126.17361	Wargul Wargul Island
R171906	WAM	TB	M	F	-16.25639	123.82444	Nw Molema Island
R171907	WAM	TB	M	F	-16.55778	123.35472	Long Island
R171908	WAM	TB	M	F	-16.41306	123.17917	Sunday Island
R171909	WAM	TB	M	F	-15.90972	124.46278	Balami ridge
R171912	WAM	TB	M	F	-16.22250	123.44972	Hidden Island
R171914	WAM	TB	M	NA	-16.55639	123.35528	Long Island
R171916	WAM	TB	M	F	-16.62361	123.47139	Lachlan Island
R171918	WAM	TB	M	M	-16.25300	123.82280	Nw Molema Island
R171919	WAM	TB	M	NA	-15.91083	124.46056	Balami ridge
R171921	WAM	TB	M	F	-15.94944	124.55944	Storr Island
R171928	WAM	TB	M	NA	-15.90944	124.46250	Balami ridge
R171930	WAM	TB	M	NA	-15.91194	124.46250	Balami ridge
R171933	WAM	TB	M	M	-15.91028	124.46139	Balami ridge
ABTC61613	SAM	TB*	SD	NA	-15.00500	129.58330	Victoria River

Table S3.2 jModelTest substitution models used with the two StarBeast2 datasets and fragment length of loci. Loci designation based on *Anolis carolinensis* genome and sequence size that was retrieved for all used samples for each locus.

Loci	Best jmodelTest model	Sequence size	Dataset
ENSACAP00000001542_exon1	K80	378	StarBeast2 gene set1
ENSACAP00000003748_exon1	HKY	624	StarBeast2 gene set1
ENSACAP00000016468_exon2	K80	516	StarBeast2 gene set1
ENSACAP00000014986_exon1	K80	588	StarBeast2 gene set1
ENSACAP00000014062_exon1	K80+G	474	StarBeast2 gene set1
ENSACAP00000016846_exon5	HKY+G	945	StarBeast2 gene set1
ENSACAP00000013421_exon9	HKY	354	StarBeast2 gene set1
ENSACAP00000006894_exon1	HKY+G	900	StarBeast2 gene set1
ENSACAP00000008371_exon1	HKY+G	318	StarBeast2 gene set1
ENSACAP00000012201_exon13	HKY+G	930	StarBeast2 gene set1
ENSACAP00000000634_exon4	K80+G	972	StarBeast2 gene set1
ENSACAP00000003365_exon1	K80+G	684	StarBeast2 gene set1
ENSACAP00000001488_exon1	K80+G	330	StarBeast2 gene set1
ENSACAP00000002281_exon14	HKY+G	612	StarBeast2 gene set1
ENSACAP00000005482_exon1	HKY+G	432	StarBeast2 gene set1
ENSACAP00000016542_exon10	HKY+G	360	StarBeast2 gene set1
ENSACAP00000004256_exon6	HKY+G	534	StarBeast2 gene set1
ENSACAP00000001948_exon6	HKY	954	StarBeast2 gene set1
ENSACAP00000003779_exon1	K80+G	672	StarBeast2 gene set1
ENSACAP00000005126_exon23	K80+G	456	StarBeast2 gene set1
ENSACAP00000003301_exon2	K80+G	366	StarBeast2 gene set2
ENSACAP00000001732_exon2	K80+G	810	StarBeast2 gene set2
ENSACAP00000004927_exon21	HKY+G	702	StarBeast2 gene set2
ENSACAP00000002002_exon1	F81	624	StarBeast2 gene set2
ENSACAP00000002929_exon1	HKY+G	480	StarBeast2 gene set2
ENSACAP00000005547_exon1	K80	438	StarBeast2 gene set2
ENSACAP00000003869_exon1	K80+G	528	StarBeast2 gene set2
ENSACAP00000011827_exon1	K80+G	504	StarBeast2 gene set2
ENSACAP00000019243_exon1	JC	480	StarBeast2 gene set2
ENSACAP00000001953_exon1	K80	624	StarBeast2 gene set2
ENSACAP00000008152_exon1	K80+G	396	StarBeast2 gene set2
ENSACAP00000012422_exon5	K80+G	588	StarBeast2 gene set2
ENSACAP00000000100_exon1	K80+G	534	StarBeast2 gene set2
ENSACAP00000009963_exon1	K80+G	522	StarBeast2 gene set2
ENSACAP00000001814_exon1	K80+G	312	StarBeast2 gene set2
ENSACAP00000012909_exon8	K80	300	StarBeast2 gene set2
ENSACAP00000011288_exon1	K80+G	564	StarBeast2 gene set2
ENSACAP00000008956_exon1	HKY+G	576	StarBeast2 gene set2
ENSACAP00000019346_exon9	K80+G	489	StarBeast2 gene set2
ENSACAP00000002448_exon12	K80+G	378	StarBeast2 gene set2

Table S3.3 Priors used for Starbeast2 and SNAPP analyses.

StarBeast2 priors			
	Lower limit	Upper limit	Distribution
Speciation rate	10	1000	1/X
Clock rates	-	-	Lognormal (real-space mean = 1, sd = 1)
Population mean	0.01	0.0001	1/X
HKY kappa (I.I.D.)	-	-	Lognormal (log-space mean = 1, sd = 1.25)
Gamma rates (I.I.D.)	-	-	Exponential (rate = 1)
SNAPP priors			
	Lower limit	Upper limit	Distribution
Speciation rate	1	100	Uniform
Population thetas (I.I.D.)	-	-	Gamma (scale = beta, shape = 2)
Beta	20	200	Uniform

Table S3.4 Summary of MANOVA results testing for significant interaction with mtDNA lineage within *C. johnstonei*. The results for testing normality and heteroscedasticity are also presented for both the log-transformed and the log and size-corrected dataset. Bold are significant p-values for the MANOVA results. After removing samples with missing data, analyses were performed with a total of 92 specimens.

	Log transformed				Log and size corrected			
	Shapiro-Wilk Test	Levene's Test	F value	p-value	Shapiro-Wilk Test	Levene's Test	F value	p-value
Snout vent length	0.48	0.06	27.47	**1.05E-06**				
Axilla to groin length	0.41	0.21	23.62	**4.93E-06**	0.38	0.50	0.05	8.21E-01
Head length	0.10	0.08	25.23	**2.56E-06**	0.35	0.27	0.83	3.66E-01
Head width	0.26	0.25	23.08	**6.18E-06**	0.32	0.96	0.32	5.71E-01
Head depth	0.06	0.18	45.57	**1.40E-09**	0.10	0.28	10.93	**1.36E-03**
Forelimb length	0.63	0.63	38.95	**1.40E-08**	0.39	0.60	7.39	**7.89E-03**
Hindlimb length	0.26	0.14	34.76	**6.42E-08**	0.11	1.00	5.16	**2.55E-02**
Nasals separation	0.16	0.15	6.04	**1.59E-02**	0.50	0.15	7.12	**9.02E-03**
Ear aperture length	0.33	0.98	20.63	**1.72E-05**	1.82E-03	0.93	2.63	1.09E-01
Palpebral distance length	0.17	0.06	10.10	**2.03E-03**	0.30	0.07	0.01	9.06E-01
Eye to ear distance	0.09	1.51E-03	32.06	**1.76E-07**	0.37	0.09	3.32	7.17E-02
MANOVA			8.37	**1.37E-09**			3.49	**7.35E-04**

Table S3.5 Summary of MANOVA results testing for significant interaction with mtDNA lineage within *C. triacantha*. The results for testing normality and heteroscedasticity are also presented for both the log-transformed and the log and size-corrected dataset. Bold are significant p-values for the MANOVA results.
* - non-normal variables with significant support with the non-parametric Wilcoxon test. After removing samples with missing data, analyses were performed with a total of 94 specimens.

	Log transformed				Log and size corrected			
	Shapiro-Wilk Test	Levene's Test	F value	p-value	Shapiro-Wilk Test	Levene's Test	F value	p-value
Snout vent length	3.75E-05	0.90	7.62	0.00696*				
Axilla to groin length	1.38E-04	0.57	5.93	**1.68E-02**	0.24	0.54	0.30	5.87E-01
Head length	3.63E-03	0.87	19.34	**2.94E-05**	0.02	0.33	13.92	**0.00033***
Head width	2.35E-03	0.76	7.89	**6.09E-03**	0.16	0.06	0.51	4.78E-01
Head depth	1.94E-05	0.96	4.10	**4.58E-02**	0.04	0.81	0.53	4.68E-01
Forelimb length	0.08	0.09	10.76	**1.46E-03**	0.24	0.61	2.80	9.75E-02
Hindlimb length	5.65E-06	0.40	6.70	**1.12E-02**	5.93E-05	5.22E-03	0.31	5.79E-01
Nasals separation	0.05	0.47	1.64	**2.04E-01**	0.91	0.65	3.23	7.55E-02
Ear aperture length	0.16	0.33	9.99	**2.14E-03**	0.32	0.56	2.56	1.13E-01
Palpebral distance length	2.58E-03	0.84	2.97	8.84E-02	5.69E-06	0.65	4.00E-03	9.48E-01
Eye to ear distance	4.10E-03	0.60	13.41	**4.18E-04**	0.03	0.83	5.03	**0.0273***
	Manova		4.92	**8.07E-06**			3.69	**2.75E-04**

Table S3.6 Summary of Generalized Linear modelling with a Poisson distribution analyses for relevant meristic variables, presenting estimates and respective confidence intervals (C.I.). Bold correspond to significant p-values. After removing samples with missing data, analyses were performed with a total of 85 and 83 specimens for *C. johnstonei* and *C. triacantha*, respectively.

		Estimate	C.I.	Std. Error	z-value	p-value
C. johnstonei	Lamellae under the 3rd finger	0.162	0.065-0.257	0.049	3.302	**9.60E-04**
	Lamellae under the 4th toe	0.142	0.059-0.225	0.042	3.370	**7.51E-04**
	Ear lobules number	0.319	0.198-0.439	0.062	5.177	**2.25E-07**
C. triacantha	Lamellae under the 3rd finger	0.020	-0.083-0.125	0.053	0.378	0.705
	Lamellae under the 4th toe	-0.006	-0.096-0.086	0.046	-0.121	0.904
	Ear lobules number	0.412	0.238-0.592	0.090	4.555	**5.24E-06**

Table S3.7 Descriptive table with measurements and meristic data for each main lineage within *C. johnstonei* and *C. triacantha*.

	Johnstonei A	Johnstonei B	Triacantha A	Triacantha B
Snout vent length	36.39 (66) (21.84 - 43.49)	41.83 (35) (27.93 - 51.44)	36.55 (31) (23.78 - 44.98)	40.07 (68) (24.72 - 49.12)
Axilla to groin length	16.31 (66) (8.53 - 23.86)	19.13 (35) (11.45 - 24.37)	16.12 (31) (9.56 - 21.53)	17.62 (68) (10.77 - 24.44)
Head length	7.63 (66) (4.88 - 9.8)	8.67 (35) (5.57 - 10.61)	7.24 (31) (5.17 - 9.14)	8.25 (68) (5.34 - 10.1)
Head width	5.45 (66) (3.7 - 7.04)	6.17 (35) (4.27 - 7.92)	5.53 (31) (4.06 - 6.89)	6 (68) (4.04 - 7.64)
Head depth	3.59 (66) (1.81 - 4.49)	4.48 (35) (2.48 - 6.07)	3.57 (31) (1.59 - 4.34)	3.81 (68) (2.4 - 4.86)
Nasals separation	2.04 (64) (1.43 - 2.46)	2.19 (33) (1.66 - 2.67)	2.16 (30) (1.67 - 2.75)	2.23 (68) (1.55 - 2.71)
Ear aperture length	1.01 (62) (0.50 - 1.44)	1.27 (32) (0.85 - 2.16)	1.13 (30) (0.64 - 1.73)	1.33 (67) (0.69 - 1.96)
Palpebral disc length	1.31 (62) (1.05 - 1.59)	1.44 (32) (1.04 - 2.13)	1.41 (30) (0.99 - 1.71)	1.46 (67) (0.92 - 1.81)
Eye to ear distance	2.66 (62) (1.79 - 3.51)	3.26 (32) (2.10 - 4.4)	2.41 (30) (1.55 - 2.90)	2.75 (67) (1.47 - 3.61)
Forelimb length	9.51 (63) (6.17 - 14.14)	11.45 (33) (8.60 - 13.94)	10.49 (29) (7.94 - 12.27)	11.53 (67) (7.27 - 14.54)
Hindlimb length	14.82 (63) (8.80 - 20.12)	17.77 (33) (12.67 - 23.18)	16.74 (29) (10.73 - 24.01)	18.08 (68) (12.37 - 21.21)
Lamellae under the 3rd finger	16.75 (63) (9 - 20)	19.69 (35) (17 - 23)	18.83 (30) (16 - 22)	19.27 (62) (11 - 24)
Lamellae under the 4th toe	22.83 (63) (15 - 27)	26.31 (35) (21 - 30)	24.83 (29) (23 - 28)	24.82 (62) (18 - 30)
Ear lobules number	9.68 (63) (5 - 16)	13.31 (35) (8 - 18)	5.86 (28) (1 - 8)	8.81 (62) (1 - 14)
Number of keels	2 (63) (2 - 2)	2.34 (35) (2 - 3)	3 (30) (3 - 3)	3 (63) (3 - 3)
Number of supraciliary scales	6.94 (63) (6 - 8)	6.94 (35) (5 - 8)	6.07 (30) (5 - 7)	5.79 (63) (5 - 7)
Number of Supralabial scales	7.03 (63) (7 - 8)	7.14 (35) (7 - 8)	7.1 (30) (7 - 8)	7 (63) (6 - 8)
Number of infralabial scales	6.27 (63) (6 - 7)	6.11 (35) (6 - 7)	6.3 (30) (6 - 7)	6.06 (63) (5 - 7)

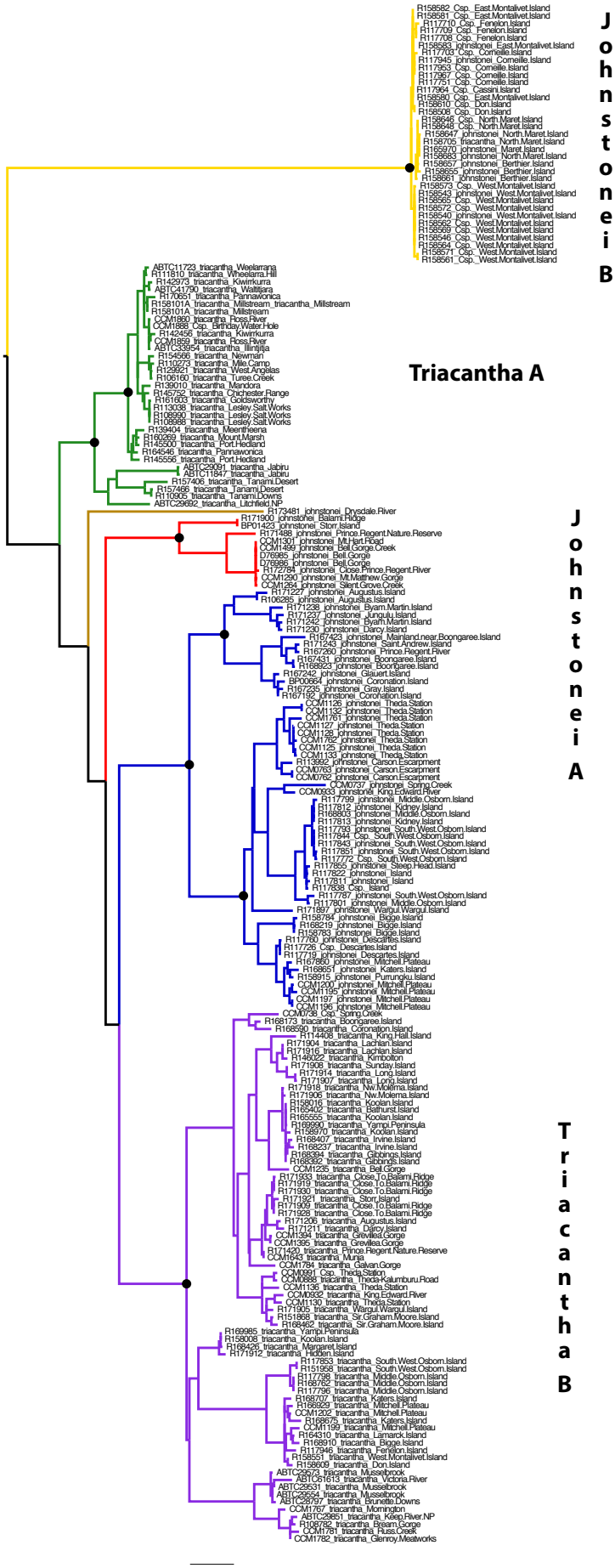


Figure S3.1 mtDNA ND4 maximum likelihood phylogenetic tree of *Carlia triacantha* and *Carlia johnstonei*, from Afonso Silva et al. (2017), with specimens that were analysed by morphological analyses. Sample label includes tissue number, original ID and sampling location.

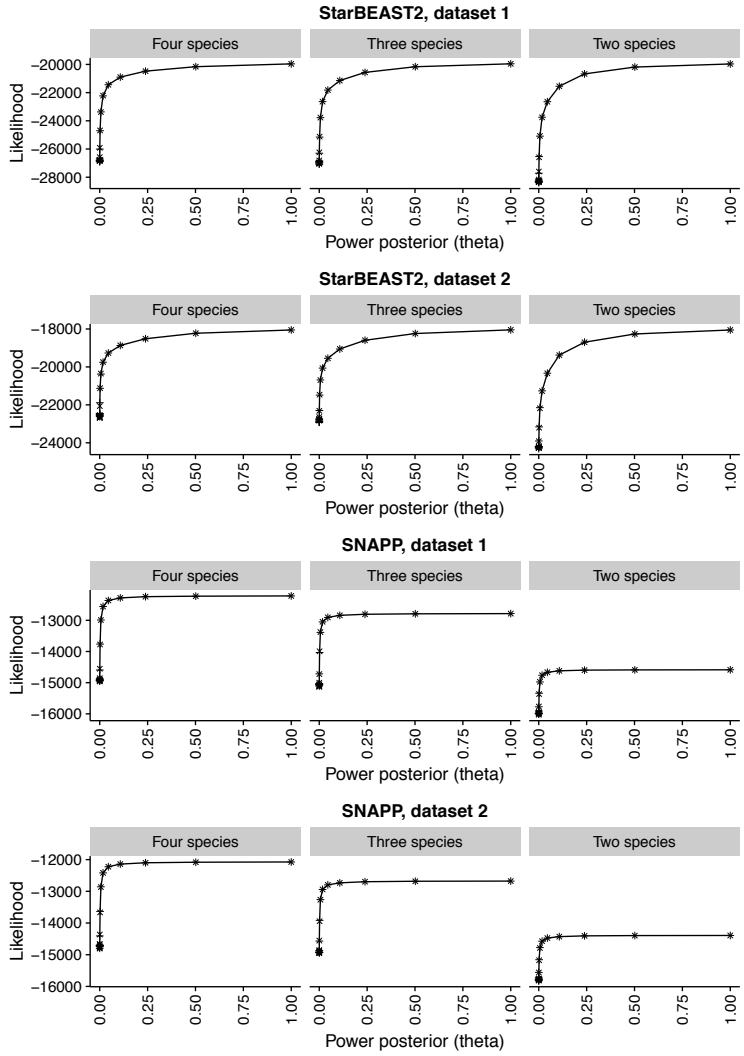


Figure S3.2 Stepping-stone computation of marginal likelihoods for Bayes factor species delimitation. This method calculates the marginal likelihoods from the area under the likelihood posterior curve. Two replicate chains were run for each method and dataset. The mean likelihoods are plotted with + symbols for one chain, and × symbols for the other. Segmented lines approximating the curve are plotted by connecting the likelihoods averaged for both chains.

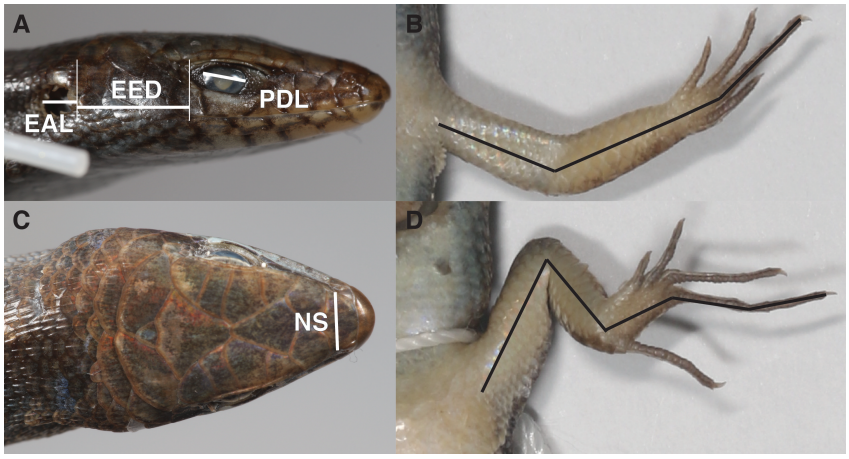


Figure S3.3 Measurements taken from photos. A) Lateral view of specimen with ear aperture length (EAL), eye to ear distance (EED) and palpebral disc length (PDL). C) Nasal separation (NS) measured in dorsal view. C) and D) correspond to forelimb (FLL) and hindlimb length (HLL) measurements in ventral view. Photos by Damien Esquerré.

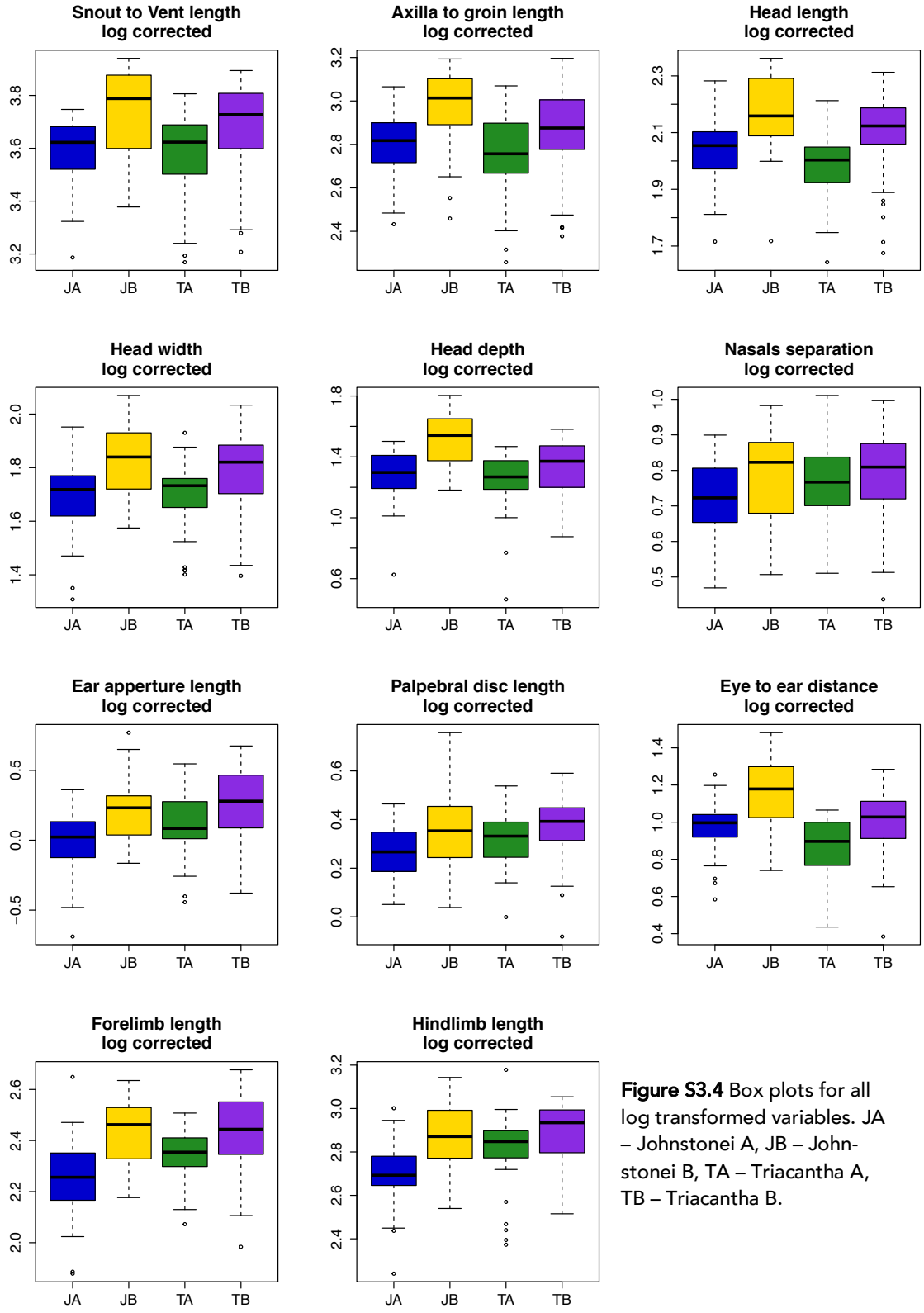


Figure S3.4 Box plots for all log transformed variables. JA – Johnstonei A, JB – Johnstonei B, TA – Triacantha A, TB – Triacantha B.

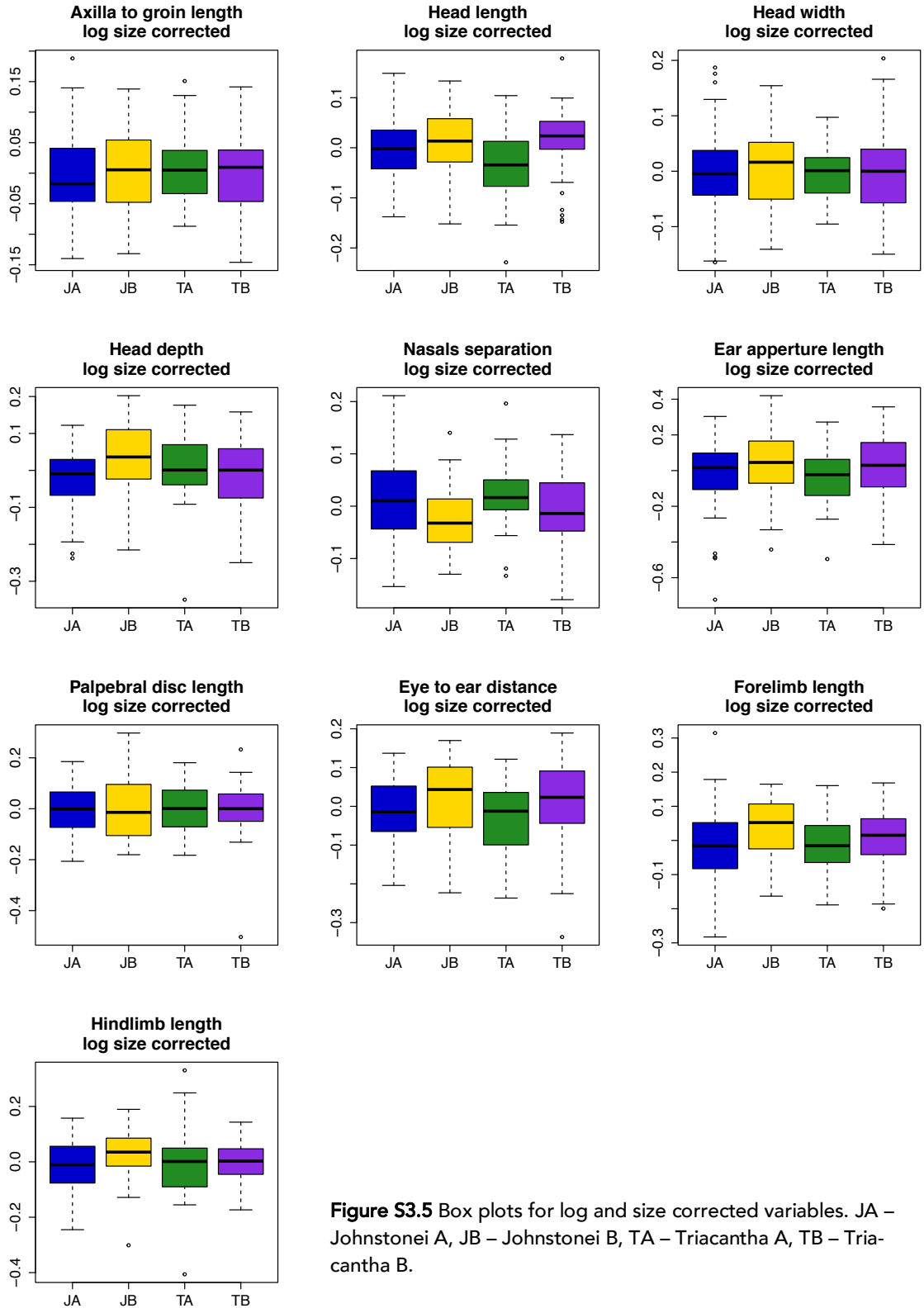


Figure S3.5 Box plots for log and size corrected variables. JA – Johnstonei A, JB – Johnstonei B, TA – Triacantha A, TB – Triacantha B.

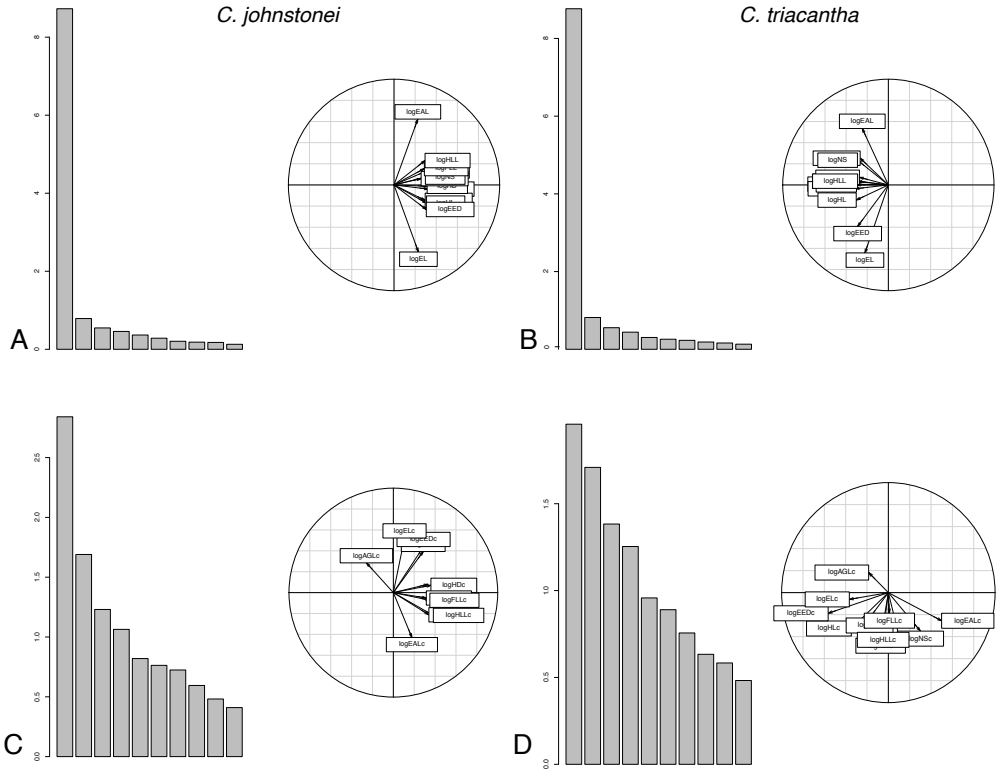


Figure S3.6 PCA loadings and variables importance of PCA with log transformed data (A, B) and with size corrected data (C, D) for *C. johnstonei* and *C. triacantha*.

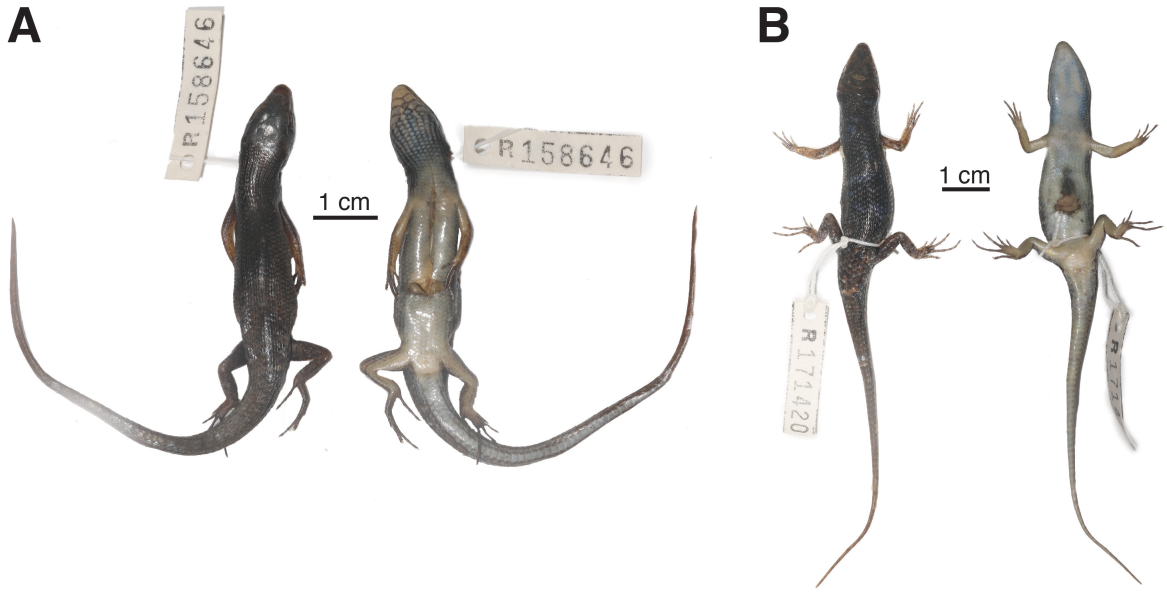


Figure S3.7 Dorsal and ventral view of holotypes of *C. insularis* sp. nov. (A), specimen WAM R158646, and *C. isostriacantha* sp. nov. (B), specimen WAM R171420. All photos by Damien Esquerré.

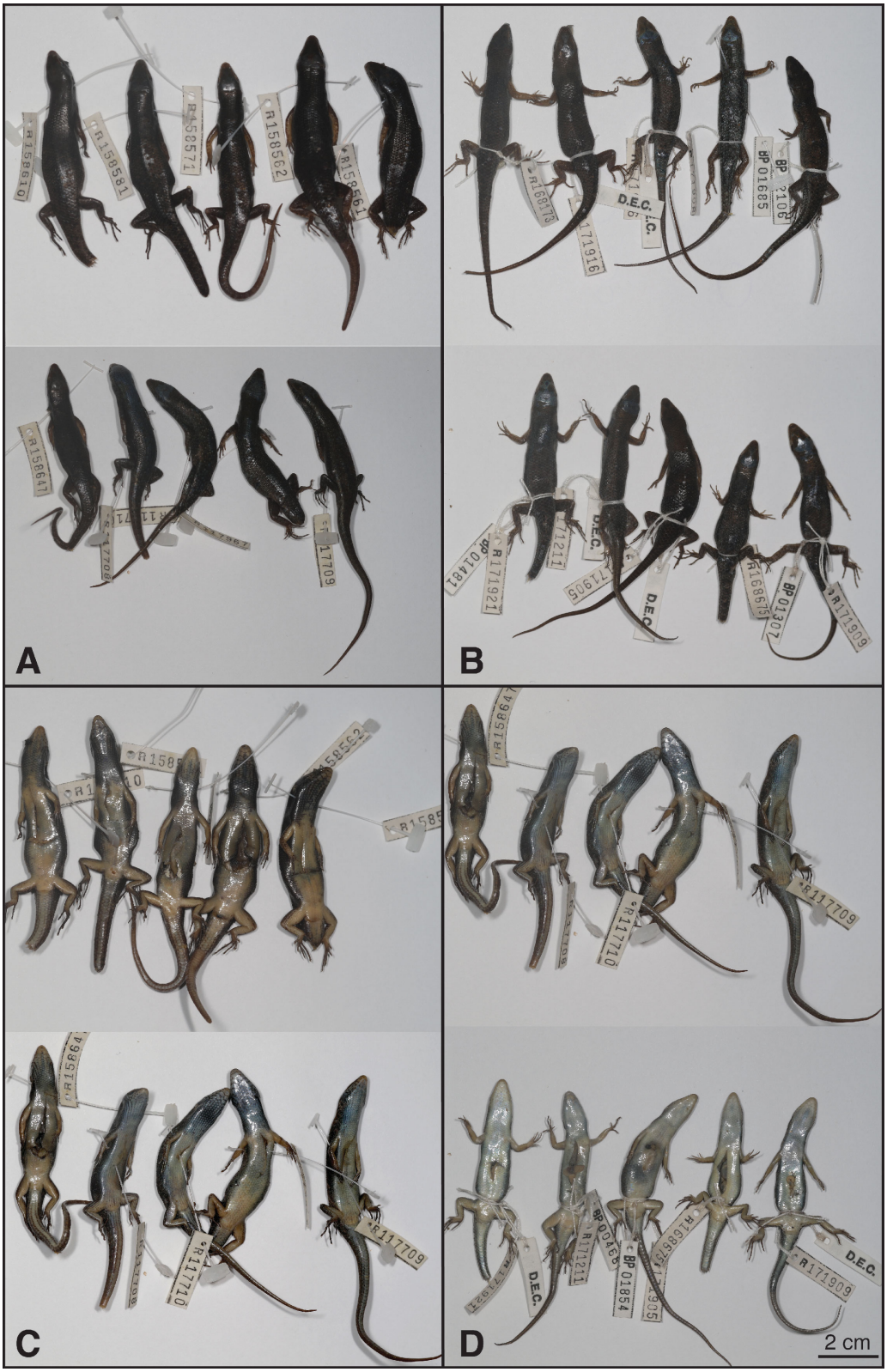


Figure S3.8 Dorsal and ventral paratype photos for *C. insularis* sp. nov (A, C) and for *C. isostriacantha* sp. nov. (B, D). All photos by Damien Esquerré.

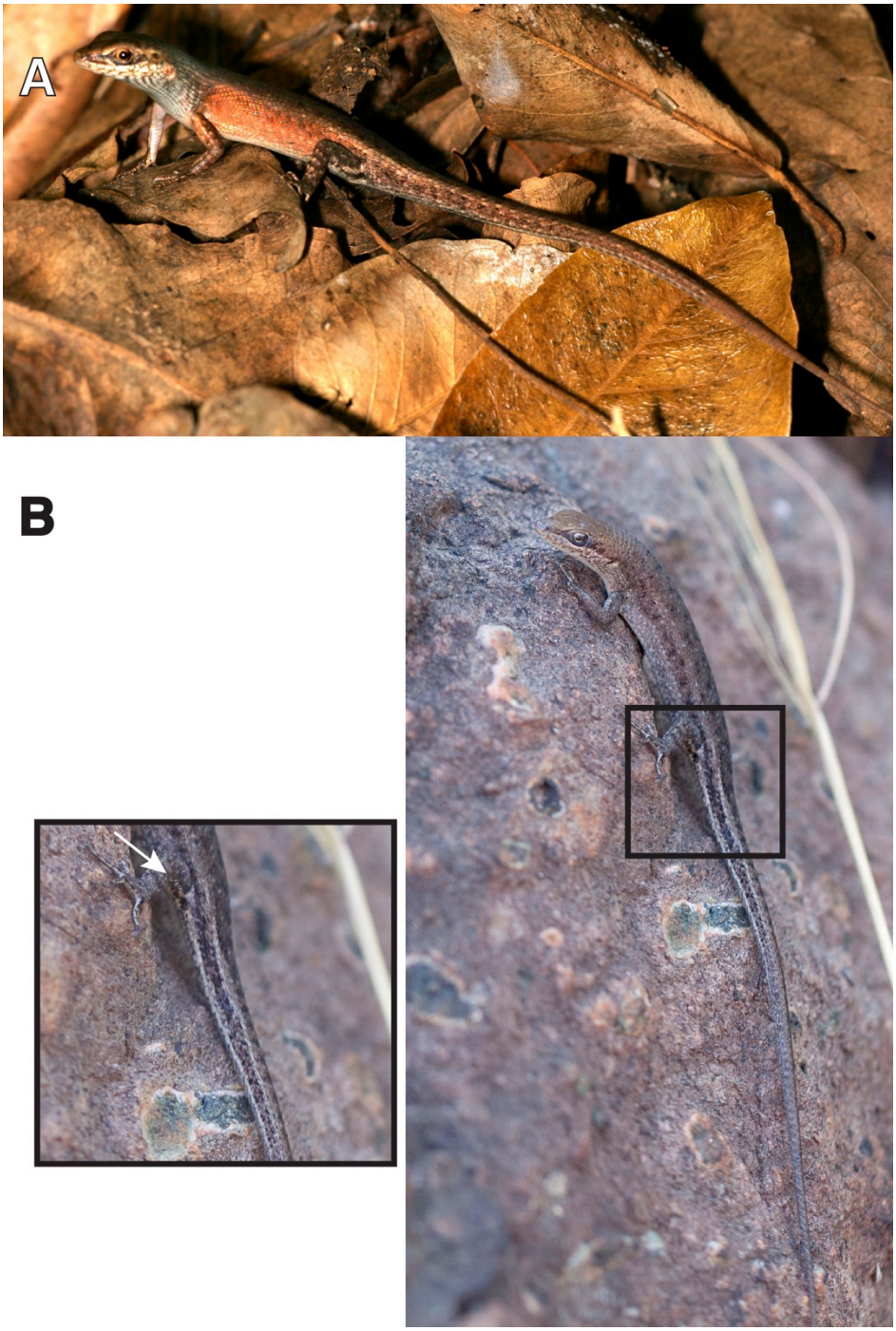


Figure S3.9 Photos with live animals showing breeding colours of *C. insularis* sp. nov. (A, photo by Russell Barrett) and a potential diagnostic trait in *C. isostriacantha* sp. nov. (B, photo by Mark Allen). The white arrow points to the potential white line trait that distinguish this species from *C. triacantha*.

Chapter IV

PHYLOGENOMICS OF A RAPID RADIATION: THE AUSTRALIAN RAINBOW SKINKS

Jason G. Bragg[1,2], Sally Potter[1], Ana C. Afonso Silva[1,3], Conrad J. Hoskin[4], Benjamin Y. H. Bai[5], Craig Moritz[1]

[1] Research School of Biology and Centre for Biodiversity Analysis, Australian National University
[2] Herbarium of NSW, Royal Botanic Gardens & Domain Trust, Sydney
[3] cE3c - Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016, Lisboa, Portugal
[4] College of Science & Engineering, James Cook University, Townsville, Qld 4811
[5] Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK

The original work described in this chapter is a collaborative study lead by Jason Bragg has been published and should be referenced as:
Bragg JG., Potter S., Afonso Silva AC., Hoskin CJ., Bai BYH., Moritz C. (2018). Phylogenomics of a rapid radiation: the Australian rainbow skinks. *BMC Evolutionary Biology*, **18**, 15.

Abstract

Background: The application of target capture with next-generation sequencing now enables phylogenomic analyses of rapidly radiating clades of species. But such analyses are complicated by extensive incomplete lineage sorting, demanding the use of methods that consider this process explicitly, such as the multispecies coalescent (MSC) model. However, the MSC makes strong assumptions about divergence history and population structure, and when using the full Bayesian implementation, current computational limits mean that relatively few loci and samples can be analysed for even modest sized radiations. We explore these issues through analyses of an extensive (>1000 loci) dataset for the Australian rainbow skinks. This group consists of 3 genera and 41 described species, which likely diversified rapidly in Australia during the mid-late Miocene to occupy rainforest, woodland, and rocky habitats with corresponding diversity of morphology and breeding colouration. Previous phylogenetic analyses of this group have revealed short inter-nodes and high discordance among loci, limiting the resolution of inferred trees. A further complication is that many species have deep phylogeographic structure – this poses the question of how to sample individuals within species for analyses using the MSC.

Results: Phylogenies obtained using concatenation and summary coalescent species tree approaches to the full dataset are well resolved with generally consistent topology, including for previously intractable relationships near the base of the clade. As expected, branch lengths at the tips are substantially overestimated using concatenation. Comparisons of different strategies for sampling haplotypes for full Bayesian MSC analyses (for one clade and using smaller sets of loci) revealed, unexpectedly, that combining haplotypes across divergent phylogeographic lineages yielded consistent species trees.

Conclusions: This study of more than 1000 loci provides a strongly-supported estimate of the phylogeny of the Australian rainbow skinks, which will inform future research on the evolution and taxonomy of this group. Our analyses suggest that species tree estimation with the MSC can be quite robust to violation of the assumption that the individuals representing a taxon are sampled from a panmictic population.

Keywords: skink, lizard, phylogenomics, multispecies coalescent, exon capture

4.1 Background

Clades that have undergone recent and rapid radiations offer unique insights into the processes that drive diversification. They showcase the effects of high rates of speciation, often driven by adaptation, and help us understand these processes in ways that are applicable more generally across the tree of life [1]. A typical starting point for making inferences about these macro-evolutionary processes is a phylogeny of species [2], describing the history of population divergences leading to the extant taxa [3]. However, it can be challenging to estimate a phylogeny for rapidly radiating clades. This is because short periods between population divergences can lead to high levels of incomplete sorting of ancestral variation (incomplete lineage sorting, ILS), which manifests in phylogenetic datasets as discordance among gene trees [4].

Phylogenomic methods have the potential to assist with these inferential challenges, by generating sequence information for hundreds or thousands of loci. In principle, these large samples of independent genealogies (gene trees) allow the estimation of a phylogeny while explicitly taking account of the genealogical discordance arising from ILS. This has been formalised in an approach called the multispecies coalescent (MSC) model. The 'full' MSC model can be used to infer a species tree and a suite of gene trees simultaneously in a Bayesian framework, and is implemented in several major packages [5,6]. Unfortunately, the full MSC is computationally intensive and, despite recent advances, its application to large datasets is not yet feasible [7,8]. A number of 'summary' MSC approaches have been developed, where gene trees are first estimated, and these gene trees are analysed to infer the species tree [9,10]. While these methods are often dramatically faster than the full MSC, and therefore permit analyses of larger datasets in practical computation times, they also often infer trees less accurately [7]. They also have other limitations, such as inferring topologies but not time-scaled branch lengths (e.g., ASTRAL, [11]).

A common alternative to the MSC is to simply infer a tree based on a concatenated alignment of sequences from multiple loci, which is computationally tractable, but can infer an incorrect topology where there is extensive discordance among genealogies due to ILS [12]. Concatenation analyses will also over-estimate branch lengths with a proportionally large effect towards the tips [7], because populations must coalesce at least as recently as their gene trees. This potentially has serious consequences for macro-evolutionary inference, particularly because branch lengths might be overestimated with greater bias near the tips, giving the appearance of reduced rates of diversification. Here we examine a phylogenomic dataset for a rapidly radiating clade of lizards. We analyse the data using multiple approaches, and compare the inferred topologies and branch lengths across different methods.

Australian skinks exemplify the opportunities and the challenges of studying clades that have diversified rapidly (e.g., [13-16]). Here we examine the rainbow skinks, a large clade that likely diversified rapidly during the mid-late Miocene. The three recognised genera of

rainbow skinks – *Carlia*, *Lygisaurus* and *Liburnascincus* – together contain more than 60 species (http://reptile-database.reptarium.cz/; accessed 2 December, 2017). A majority of these species occur in Australia, but at least 23 recognised species (20 *Carlia* and 3 *Lygisaurus*) occur in New Guinea and Wallacea [17-20]. Resolving the number and boundaries of species in New Guinea (e.g., within the diverse *fusca* group; E. Rittmeyer, unpublished data) remains a work in progress. Therefore we focus on the Australian representatives of this clade, as for these we have a relatively robust taxonomy and extensive distributional and phylogeographic data, as well as prior phylogenetic hypotheses. The 26 recognised Australian *Carlia* species [21] are distributed across northern and eastern Australia. They inhabit leaf litter and rocky areas in diverse vegetation types from rainforest to arid areas, with the highest diversity in dry woodland habitats in tropical Australia [18]. The 11 Australian species of *Lygisaurus* occur in mesic parts of the north and east, in rainforest, woodland and rocky habitats [17]. *Liburnascincus* consists of four species that are all restricted to rocky habitats in north-eastern Australia and live a truly saxicoline lifestyle [20]. The rainbow skinks reach their highest diversity in north-east Australia, with sites that have up to 10 species occurring in close proximity.

Previous studies of the rainbow skinks have linked their habitat affinities to aspects of their morphology [22-24]; for example, *Liburnascincus* have the long-limbed morphology typical of lizards in rocky habitats [23]. The *Carlia* species have also diversified in breeding colours, with males having spectacular throat and flank breeding colours that differ among species and show some relationship to shifts to more open habitats [24]. However, efforts to understand the diversification of the rainbow skinks, including macro-evolutionary analyses of phenotypes in relation to habitat shifts, have been hindered by the difficulty of estimating a robust species tree. Studies based on mitochondrial gene sequences [25] and small numbers of nuclear loci [26] did not resolve relationships deep in the tree, including among the genera, and found discordance among loci [26]. These observations are consistent with a rapid diversification, and high levels of ILS. Resolving the phylogeny of rainbow skinks has been exacerbated by phenotypically cryptic divergence in some clades, manifesting as deeply divergent and sometimes paraphyletic phylogeographic lineages within described species [27-31], which can be strongly isolated when in secondary contact [32]. Some of these complexes have been revised taxonomically (e.g., *C. fusca* group [33]; C. pectoralis group [34]; *C. triacantha* group, in part [21]) but this remains to be done for other species with deep phylogeographic structure (e.g., [31]).

In this study, we explore the two major challenges in estimating the evolutionary history of rainbow skinks, both common in rapid radiations: gene tree discordance at short inter-nodes near the base of the tree, and cryptic diversity near the tips of the tree. We use exon capture sequencing [35-37] to generate a large multi-locus dataset for representatives of all named and current candidate species of the Australian rainbow skinks. We infer the relationships among species using a variety of different phylogenetic methods, including full

MSC (StarBEAST2, [8]) and summary MSC (ASTRAL-II, [11]) approaches, as well as maximum likelihood estimation of concatenated alignments.

We also exploit well-characterised intra-specific lineage diversity in several focal species of *Carlia* [29,31] to investigate the consequences of violating an assumption of MSC estimation: that the alleles representing each 'taxon' are drawn from a panmictic population. In many phylogenetic studies, such as ours, a modest number of samples are collected at different localities, and are aggregated to represent a 'taxon' when estimating the species tree. In the absence of population-level sampling, it can be difficult to determine when samples are representatives of the same panmictic population, demes that are connected by gene flow, or long-isolated lineages. To explore the consequences of systematic deviation from this assumption, we estimated species trees for a set of four closely related species of *Carlia* skinks with known, deep phylogeographic structure, drawing haplotypes for each species (i) from the same or (ii) separate individuals within an intraspecific lineage, and (iii) from members of divergent lineages within species.

In sum, we generate a phylogenomic dataset for the Australian rainbow skinks – a clade that has undergone a rapid radiation – and use this dataset to examine the consequences of different approaches to species tree analysis, including different methods for tree estimation, and different ways of aggregating samples into 'taxa' for analysis.

4.2 Methods

4.2.1 Sampling

We obtained tissue samples for all but one of the 41 currently named species of Australian *Carlia*, *Lygisaurus*, and *Liburnascincus*, (Table S4.1), the exception being *Lygisaurus abscondita*, for which no tissue was available. This includes five named taxa, *Carlia insularis*, *Carlia isostriacantha*, *Carlia rimula*, *Carlia wundalthini* and *Lygisaurus rococo*, which have not previously appeared in a molecular phylogeny. Two of these taxa, *C. insularis* and *C. isostriacantha*, were known previously as divergent lineages of *C. johnstonei* and *C. triacantha* (respectively) [30], and described recently as separate species [21]. Where possible, we obtained at least two samples per species. For five species of *Carlia* (*C. amax*, *C. gracilis*, *C. munda*, *C. rubrigularis*, and *C. rufilatus*), we used tissue samples from multiple individuals belonging to each of two or more known phylogeographic lineages [27,29,31]. We also included representatives of additional candidate species, including samples of a divergent lineage similar to *Lygisaurus macfarlani* (from the Northern Territory), and two lineages of *C. rubrigularis* (northern and southern) [27]. We obtained a sample of an undescribed *C. fusca* group species from Waro, Papua New Guinea, because several Australian species have been placed in this group, which is predominantly Papuan [19]. Similarly, we obtained samples for a *Lygisaurus* lineage (cf. curtus) from Papua New Guinea, to provide context for our understanding of the evolutionary history of the Australian *Lygisaurus* taxa. Finally, three outgroup species were included, *Lampropholis coggeri, Lampropholis guichenoti* and *Pygmaeascincus*

timlowi, each represented by two samples. In total, our dataset included 123 samples from 46 'taxa', including 43 recognized species (Table S4.1).

4.2.2 Exon capture sequencing

For each sample, we performed exon capture sequencing. This method uses hybridization to a set of oligonucleotide probes to produce DNA libraries that are selectively enriched with a set of 'target' loci. We followed methods described by Bragg et al. [37], and the same probe set, which targets 3320 protein coding exon sequences (see Dryad repository doi:10.5061/dryad.34274 for target sequences).

We extracted DNA from each tissue sample using the salting-out method [38]. We prepared barcoded sequencing libraries for each sample using a protocol described by Meyer and Kircher [39], with modifications described by Bi et al. [36]. The barcoded DNA was pooled, along with other skink samples, in groups of 56 at a time, at equimolar concentrations, and hybridized with the probe kit (SeqCap EZ Developer Library; NimbleGen). Hybridization was performed following the manufacturer protocol, except the hybridization mix was modified slightly to contain: 1.2 µg of the pooled DNA, 5 µg of skink Cot-1 DNA (made using a method described by Trifonov et al. [40], with a sample of Lampropholis coggeri), and a set of 56 blocking oligos (1000 pmol). We performed Polymerase Chain Reactions (PCRs) to enrich the post-capture libraries (17 cycles).

We quantitated DNA in pooled pre- and post-capture DNA libraries (Bioanalyzer) and used qPCR to test for enrichment of target DNA, and de-enrichment of non-target DNA (following Bi et al. [36]). We then sequenced the post-capture libraries using an Illumina HiSeq (100-bp paired-end) instrument (Biomolecular Resource Facility, Australian National University). Raw sequencing reads are in the NCBI short read archive (BioProject PRJNA289283, see Table S4.1 for BioSample numbers).

Raw sequencing reads were cleaned using a workflow that consisted of removing duplicates, merging overlapping reads, and trimming poor quality bases and adaptor sequences. This workflow is described in detail by Singhal [41], and the code used in the present study is archived in Dryad (doi.org/10.5061/dryad.v1d32). For each sample, we then assembled the clean reads using an 'exon-specific' approach described by Bragg et al. [37]. This workflow begins by identifying sequencing reads that are homologous to each exon (blastall 2.2.26 [42], program = blastx, expectation value = 1E-9), and performs an assembly of this small subset of reads (velvet version 1.2.08 [43] assemblies with K= 31, 41, 51, 61 and 71 were merged using cap3 version 08/06/13 [44], with parameters: -o 20 –p 99). These contigs were trimmed to the exon sequence (flanking introns removed; exonerate 2.2.0 [45]). If more than one contig was assembled for a particular target locus, the best hit to the target was identified by a reciprocal best BLAST hit criterion (blastall 2.2.26 [42]). Using these assembled target sequences for each sample as a reference, the clean sequencing reads were then mapped (bowtie2, version 2.2.4 [46]), heterozygous sites identified, and overlapping reads were used to phase heterozygous sites within target loci (Genome Analysis Toolkit,

version 3.3-0-g37228af [47]). This meant that alleles at heterozygous sites could be assigned to one of two haplotype sequences (h0 and h1), though where this was not possible, the reference allele was assigned to h0 and the alternate allele to h1. Finally, the workflow produced two haplotype sequences for each locus in each sample, replacing all sites with a genotype quality score (GQ) less than 20 with an N (unknown base). The code used to call compiled software and perform other tasks in sequence assembly, mapping, calling and phasing of heterozygous sites, and creation of haplotype sequence files, is archived in Dryad (doi.org/10.5061/dryad.v1d32).

4.2.3 Alignments

For each locus, we performed alignment using MACSE (v1.01b [48]). Codons were removed from alignments if they contained a site with greater than 20% missing data (trimAl v1.4.rev15 [49]). We then estimated a gene tree using RAxML (RAxML 8.2.3 [50], -m GTRGAMMA –N20), performed 100 bootstrap replicates, and used these to calculate a relative tree certainty (TC) score [51]. TC provides an index of the information content of different loci based on agreement among bootstrap replicates. Among loci, we observed a positive association between TC score and the length of the locus alignment, and with the number of parsimony informative characters (Fig. S4.3). For the purposes of analysis, we identified two sets of loci. The first set consisted of 1384 loci containing near complete data (hereafter, complete loci, or "CL"). These loci had a sequence for every sample, and each sequence was missing data for less than 35% of sites. The second set of 304 loci was a subset of the complete loci, but were also the most informative on the basis of their TC score (TC > 0.25; hereafter, complete and informative loci, or "CIL"). We did this because several recent studies have highlighted the importance of informative gene trees in species tree estimation [16,52,53].

4.2.4 Estimating the phylogeny of the Australian rainbow skinks

In addition to exploring analytical issues, a major goal was to infer the phylogenetic tree for the Australian rainbow skinks and hence enable future comparative studies. We did this using two kinds of approaches: maximum-likelihood estimation of a tree using a concatenated alignment of multiple loci, and species tree estimation with the MSC (see Figure S4.1 for a summary of inferential analyses).

We estimated maximum likelihood trees for concatenated alignments of the CL and CIL data using IQTREE (1.3.5 [54]). In each case, we used the best substitution model (-m TEST) based on a comparison of AICc values (calculated by IQTREE). Support for the estimated topology was inferred with rapid bootstrapping ([55], parameter: -bb 1000). For the CIL alignment, we performed a second analysis, after partitioning by locus and codon position. Here, we placed the sites at codon positions 1 and 2 of each exon in a partition, and the sites at codon position 3 in another partition, and used the hclust algorithm (described by [56], implemented in IQTREE 1.3.5 [54]) to find a 'best-fit' partition model via sequential

merging of partitions. We did not perform a similar analysis for the CL dataset, which would have produced a very large number of possible partitions. To check that our inferred tree was supported by different implementations of maximum likelihood (identified as an issue in [53]), we also estimated trees for the CL and CIL alignments using RAxML 8.2.3 [50] (parameters -m GTRGAMMA –N20).

We applied the MSC to the rainbow skink data in two ways. First, we used a 'summary' coalescent method, ASTRAL-II (4.7.9 [11]), which estimates a species tree based on previously inferred gene trees for many individual loci. We began by estimating an ASTRAL species tree using the 1384 CL gene trees estimated with RAxML (8.2.3 [50], -m GTRGAMMA –N20). Node support for this tree was inferred with 100 multi-locus bootstrap replicates. However, we wanted to check that our results were robust to a range of factors that can influence phylogenetic estimation, including the implementation used for maximum likelihood estimation [53], model violation at third codon positions [57], and variation in the information content of the individual loci [16,53]. Therefore, in addition to the RAxML gene trees (approach "R", Fig. S4.1), we estimated gene trees for the CL loci using two approaches. First, we estimated gene trees using IQTREE (-m TEST, best model selected based on AICc; approach "Q"). Second, we generated new alignments for the CL loci with third codon positions removed, and estimated gene trees for these alignments using IQTREE (-m TEST, best model selected based on AICc; approach "Q_{12}"). We estimated an ASTRAL species tree using the gene trees generated with each of the tree approaches (R, Q and Q_{12}). We also estimated ASTRAL species trees using gene trees (approaches R, Q and Q_{12}) for the CIL subset of loci. These analyses are summarised in Fig. S4.1.

Finally, we used StarBEAST2 (version 0.6.1 [8]), an implementation of the 'full' MSC, to jointly infer the species tree of the rainbow skinks along with a set of gene trees, in a Bayesian Markov Chain Monte Carlo framework. This approach is very computationally intensive, such that it was only possible to run for relatively small numbers of loci. With this dataset, we had difficulty obtaining convergence for StarBEAST2 chains when using more than 32 loci. We therefore performed an ensemble of nine separate StarBEAST2 runs, using different sets of 32 loci that were selected randomly (without replacement) from the CIL alignments. Each StarBEAST2 [8] run used analytical population size integration, gene tree relaxed clocks, an HKY + Γ nucleotide substitution model (with a single κ value and a single α value shared across all the data, and four gamma rate categories), and a chain length of 2^{27} states (sampled every 2^{15} states) (for details of implementation, see example xml in dryad repository doi.org/10.5061/dryad.v1d32).

For trees inferred with MSC methods, Table S4.1 indicates how samples were assigned to the 46 'taxa.' Note that *C. triacantha* and *C. isostriacantha* were treated as a single 'taxon' (for MSC analyses), prior to the recent description of *C. isostriacantha* [21]. Molecular divergence between these two taxa is low, and similar to that among lineages within other species as currently recognised. As such, we do not expect this grouping to unduly distort species tree analyses (as also demonstrated below). By contrast the newly described *Carlia insularis*

[21] was treated as a taxon separate from its sister, *C. johnstonei*, as there is much deeper molecular divergence between these two taxa [30].

We estimated trees for the rainbow skinks using a variety of different approaches, and we compared these trees in two ways. First, we compared topologies by calculating Robinson-Foulds distances (function 'RF.dist' in the R [58] package phangorn [59]). Second, we compared the branch lengths estimated by concatenation and MSC methods. We assessed this using an MSC tree whose topology was estimated using ASTRAL, with branch lengths set to the average ancestor heights from the nine StarBEAST2 posterior distributions. We then calculated the depth of each node in this tree relative to the depth of the base. We compared this to the concatenation tree that was estimated with the CL dataset, after performing a similar normalization. We forced the concatenation tree to be ultrametric (using the function 'chronos' in the R [58] package ape [60]), and calculated the depth of each node relative to the base. We note that concatenation trees (where each tip is a sample) contain more tips than species trees (where each tip is a 'taxon'). To make comparison of trees meaningful, tips were dropped from concatenation trees so that each 'taxon' (as used in MSC analyses) was represented by one randomly-selected sample.

4.2.5 Assigning alleles to taxa in the multispecies coalescent

To investigate how violation of the assumption of panmictic 'taxa' might affect MSC estimation of the species tree, we identified a 'focal' dataset, consisting of a clade of nine taxa that contained several species with well-characterized, deeply divergent lineages (see Table S4.1). We chose samples from two diverged lineages from four species (*Carlia* amax [29]; *C. munda*, *C. rufilatus* and *C. gracilis* [31]) to use in an experiment (listed in Table S4.1, focal clade lineage experiment). Here we used the full MSC to estimate a species tree for the clade using two alleles from each taxon, where these alleles were from: (i) a single individual (the two haplotypes from a single diploid; treatment "1S" or one sample), (ii) different individuals from the same lineage (treatment "2S1L," or 2 samples, from 1 lineage), and (iii) individuals from different lineages (treatment "2S2L," or 2 samples from 2 lineages). These treatments were chosen to reflect different kinds of violations of the assumptions that a taxon represents an undivided population. In the 1S treatment, the two haplotypes from a single individual would reflect genetic variation in a population in the unlikely event mating is random. The 2S2L treatment, where samples are drawn from different lineages, is a clear and deliberate violation of the assumption that the samples are drawn from the same population. We expect that the 2S1L treatment is a closer fit (relative to 1S and 2S1L) to the assumptions of the MSC, allowing somewhat for isolation by distance. For each treatment, we performed nine replicate MSC analyses, each used a set of 32 loci that were randomly selected (without replacement) from among the CL alignments, and that were used for the StarBEAST2 analyses (implemented as described above for the whole clade). For each replicate, and each treatment, we chose the two alleles used in the analysis at random. That is,

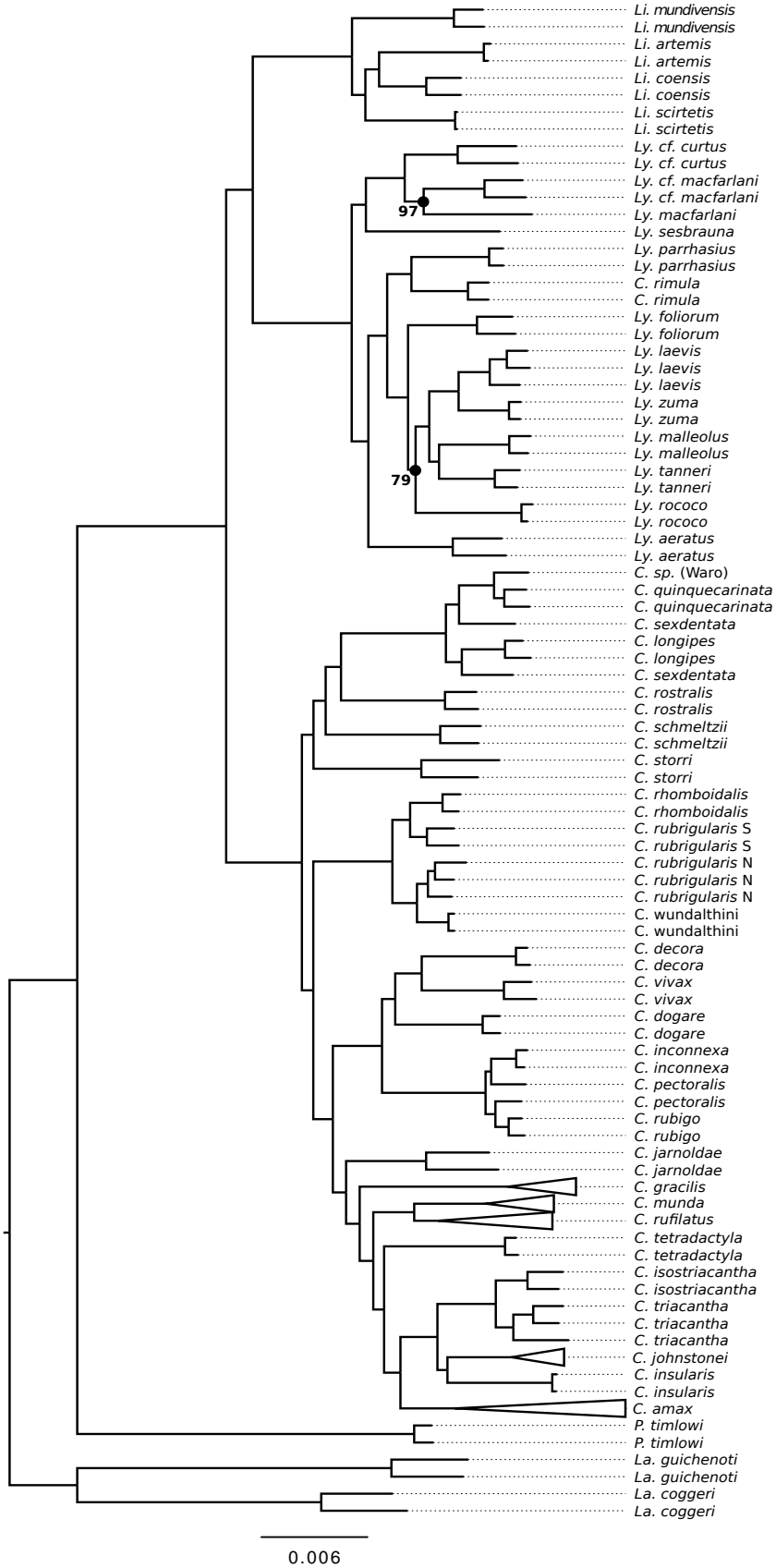


Figure 4.1 Phylogeny of the Australian rainbow skinks based on a concatenated alignment of 1384 exon sequences. This tree was estimated by maximum likelihood (implemented in IQTREE), and node labels indicate the percentage of bootstrap support (where less than 100). Clades are represented as triangles where a 'taxon' had more than three samples.

for treatment 1S, we chose an individual at random from among the set of candidates (indicated in Table S4.1), and used both haplotypes from that sample in the MSC analysis. For 2S1L, we selected a lineage at random to represent each taxon, and used an allele from each of two individuals from that lineage. For the 2S2L treatment, we used an allele from one randomly chosen individual from each of two diverged lineages to represent each taxon in the MSC analysis.

4.3 Results

We begin with a brief summary of sequencing quality and coverage statistics (Fig. S4.2). We then describe the phylogeny we inferred for the Australian rainbow skinks, with a focus on agreement and conflict among trees estimated using different approaches. We next present the results of experiments that show how the assignment of alleles to taxa can affect phylogenetic estimation under the MSC.

4.3.1 Exon capture sequencing

Across the 123 samples, we assembled sequences for an average of 2606 loci, and a minimum of 2364 loci. The sequencing depth of coverage for these loci was high, with an average of 131.4x (see Fig. S4.2). In sum, we performed high-coverage sequencing of thousands of loci, applied stringent thresholds for data completeness (resulting in the CL set, described in Methods), and performed subsequent analyses using high quality sequence data.

4.3.2 Phylogenetic estimation using concatenation and species tree methods

We estimated the genetic relationships among members of the Australian rainbow skinks using a variety of approaches (see Methods, Fig. S4.1), including concatenation (Fig. 4.1) and multispecies coalescent (MSC) analyses (Fig. 4.2). The trees inferred using different approaches were often concordant, especially at nodes that were estimated with high levels of bootstrap support. Our topology also largely supports those clades that were inferred with high confidence in a previous study [26], including monophyly of the rainbow skink genera, though with interesting exceptions, which are discussed below.

The trees we inferred using concatenated alignments received high levels of bootstrap support at almost all nodes (Fig. 4.1). In all but three cases (*C. sexdentata*, *C. rubrigularis* and *C. pectoralis*), conspecific samples form clades, consistent with the assignments of samples to taxa used in species tree analyses (Table S4.1). When we compare trees that were estimated using the concatenated CL versus concatenated CIL datasets, or used partitioned versus un-partitioned alignments (CIL), the topologies were highly concordant (i.e., low RF distances), except at a small number of nodes (Fig. 4.3).

The ASTRAL analyses also produced trees with strong bootstrap support (Fig. 4.2). Again, there was little difference in the topologies of the trees we inferred using the CL versus CIL datasets, or when we estimated the gene trees excluding third codon positions

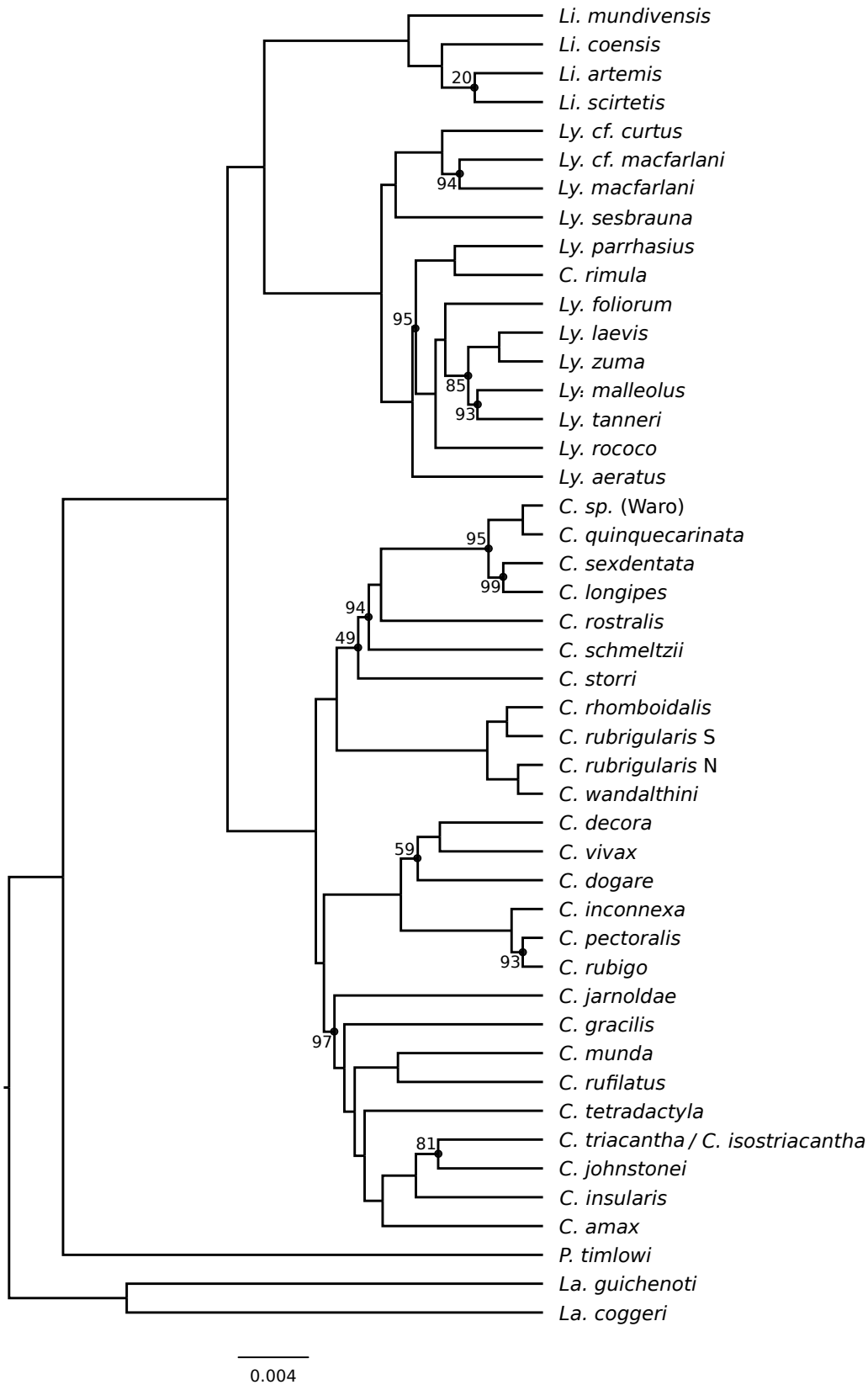


Figure 4.2 Phylogeny of the Australian rainbow skinks estimated using the multispecies coalescent (MSC). This topology was inferred using ASTRAL, based on gene trees for 304 loci (the CIL set, see text), and node labels indicate the percentage of bootstrap support (where less than 100). The branch lengths of this tree were calculated using mean ancestor heights from nine StarBEAST2 analyses, each based on a different set of 32 loci.

(Fig. 4.3). Overall, there was high concordance between the trees estimated using concatenated alignments, and those estimated using ASTRAL (Fig. 4.3).

However, the nine species trees we inferred using StarBEAST2, each based on a different set of 32 loci, often had topologies that conflicted with those estimated using ASTRAL, and with each other (Fig. 4.3). It is possible there is greater discordance among the StarBEAST2 analyses because each considers a much smaller sample of different genealogies (32 loci) than the concatenation and summary coalescent methods.

We also compared the branch lengths estimated by concatenation (shown in Fig. 4.1) and MSC (shown in Fig. 4.2) methods. In Fig. 4.4, we have plotted the relationships between the node depths estimated by the MSC (horizontal axis) and concatenation (vertical axis). We observe that many of the values lie above the line (see Figs 4.4 a and b), meaning that these node depths are overestimated (relative to the base of the tree) by concatenation. This is particularly common for nodes near the tips of the tree (Fig. 4.4b).

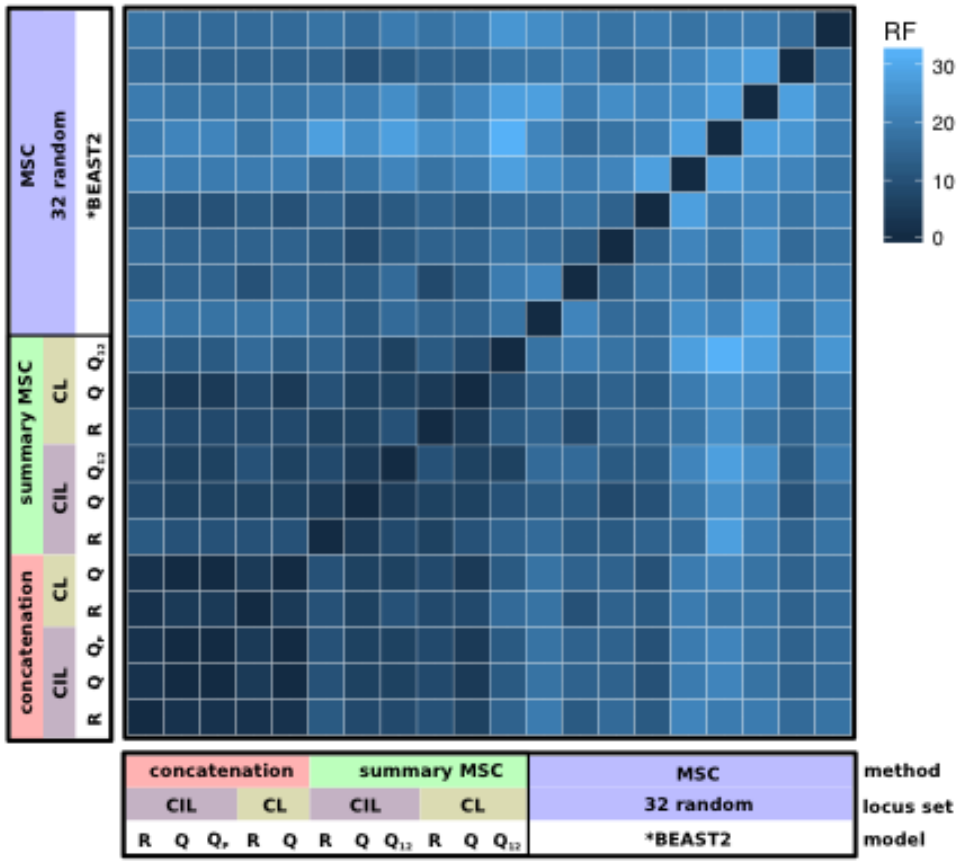


Figure 4.3 Topological discordance between phylogenies estimated using different methods. Most broadly, the different methods included concatenation, a summary MSC approach (ASTRAL) and the full MSC (StarBEAST2). We also performed analyses using different maximum likelihood implementations (RAxML, R; IQTREE, Q), partitioning (Q_P), and analyses that excluded third codon positions (Q_{12}) (see Fig. S4.1). In this matrix, the color of each cell indicates the Robinson-Foulds distance between trees estimated using two different methods, or in the case of the StarBEAST2 trees, between one of the nine replicate analyses.

4.3.3 Assignment of alleles to taxa and the MSC

In the experiment comparing sampling strategies (1S1L, 2S1L and 2S2L) there was variation in maximum clade credibility topologies among the 27 MSC analyses (Fig. 4.5b). However, the trees inferred across the nine replicates using the 2S2L treatment were more similar topologically to each other, and to the ASTRAL and concatenation trees, than were the trees inferred across replicates using the other two sampling approaches (Fig. 4.5). Additionally, the trees estimated by StarBEAST2 for the 2S2L treatment tended to have greater values for the N_e parameter than the other two treatments, but did not have greater tree heights (Fig. 4.5c).

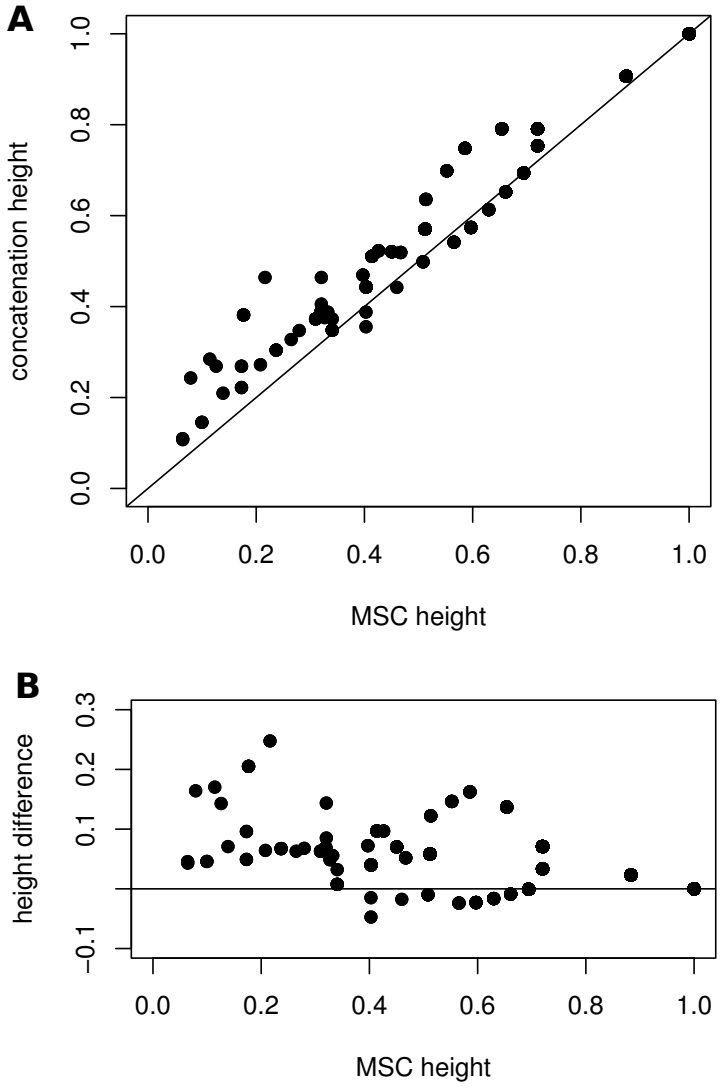


Figure 4.4 The relationship between node depths (relative to the base node) for species trees that were estimated using the multispecies coalescent (MSC) a concatenated alignment (inferred by maximum likelihood). In (a) the bivariate relationship between the inferred depths is shown, and in (b) the difference between the height inferred by concatenation and the MSC is shown, as a proportion of the depth inferred by the MSC [(concatentation – MSC) / MSC].

4.4 Discussion

4.4.1 Comparing approaches of phylogenetic estimation

When we estimated the phylogeny of the Australian rainbow skinks using different approaches, the topology was affected little by several factors that have been important in other studies, such as the implementation used for maximum likelihood estimation [53], and

the removal of third codon positions prior to analysis [57]. The topologies we inferred using concatenation and summary coalescent methods were also similar. While our results were largely robust to these factors, we note that this might not be true in general – for example in analyses of deeper relationships where mutational biases are more complex, or among closely related taxa where there is substantial introgression [61].

We observed greater discordance in topologies estimated with StarBEAST2, possibly because each of our StarBEAST2 analyses (necessarily) used fewer loci. This observation is consistent with the findings of Blom et al. [16], who analysed exon sequence data for a different, but also rapidly radiating, clade of Australian skinks, and found that species trees inferred using fewer than 80 loci were often discordant, but when a greater number of informative loci were used in analyses, species trees tended to converge on the same topology. With ongoing improvements in computational efficiency of StarBEAST2 [8], it should become possible to revisit full Bayesian species tree analysis for the taxa examined here.
Finally, our study corroborates previous observations that concatenation analyses overestimate the inferred depths of nodes near the tips of trees [7]. While this result was expected, it is useful to document this pattern, because it might have substantial consequences for downstream applications, such as macro-evolutionary analyses that examine the rate of divergence events through time.

4.4.2 Assignment of alleles to taxa and the MSC

We found it interesting that trees generated using the 2S2L sampling strategy treatment were relatively similar to each other and to the ASTRAL and concatenation trees, given that this treatment grossly and intentionally violates the panmixia assumption of the MSC. It is possible that the co-estimation of effective population size (Ne), as implemented in StarBEAST2, is adjusting appropriately for the extra diversity within taxa that are represented by alleles from different lineages. This possibility is supported by the observation of greater observed Ne values for the 2S2L treatment. In sum, these analyses show that the sampling of alleles for different taxa has consequences for trees and other parameters that are estimated with the MSC. However, the results do not add support to the concern that motivated our analysis, which was that deep intraspecific lineage diversity, and violation of relevant assumptions, might compromise phylogenetic inference. We suggest it would be worthwhile to further explore this question through a simulation study. Here datasets might be simulated for clades with known (prescribed) species trees, and where the 'species' contain lineages with different and contrasting divergence times. The simulated datasets could then be sampled, to test whether different sampling strategies were more likely to recover the known, true, tree.

4.4.3 Phylogeny and taxonomy of the Australian rainbow skinks

In the phylogeny that we estimated for the rainbow skinks, each of the genera – *Carlia*, *Lygisaurus* and *Liburnascincus* – was monophyletic. There was an exception, in that *Carlia*

rimula was placed in a clade with all members of *Lygisaurus*, as sister to *Ly. parrhasius* (Figs. 4.1 and 4.2). The morphological similarity between *Carlia rimula* and *Lygisaurus parrhasius* was noted previously in a phylogenetic study [26] (see also [19]), but genetic data were not then available for *Carlia rimula*. Based on our comprehensive and well-resolved phylogeny, we reassign *C. rimula* (Ingram and Covacevich 1980) [62] to the genus *Lygisaurus*, forming the new combination *Lygisaurus rimula* comb. nov..

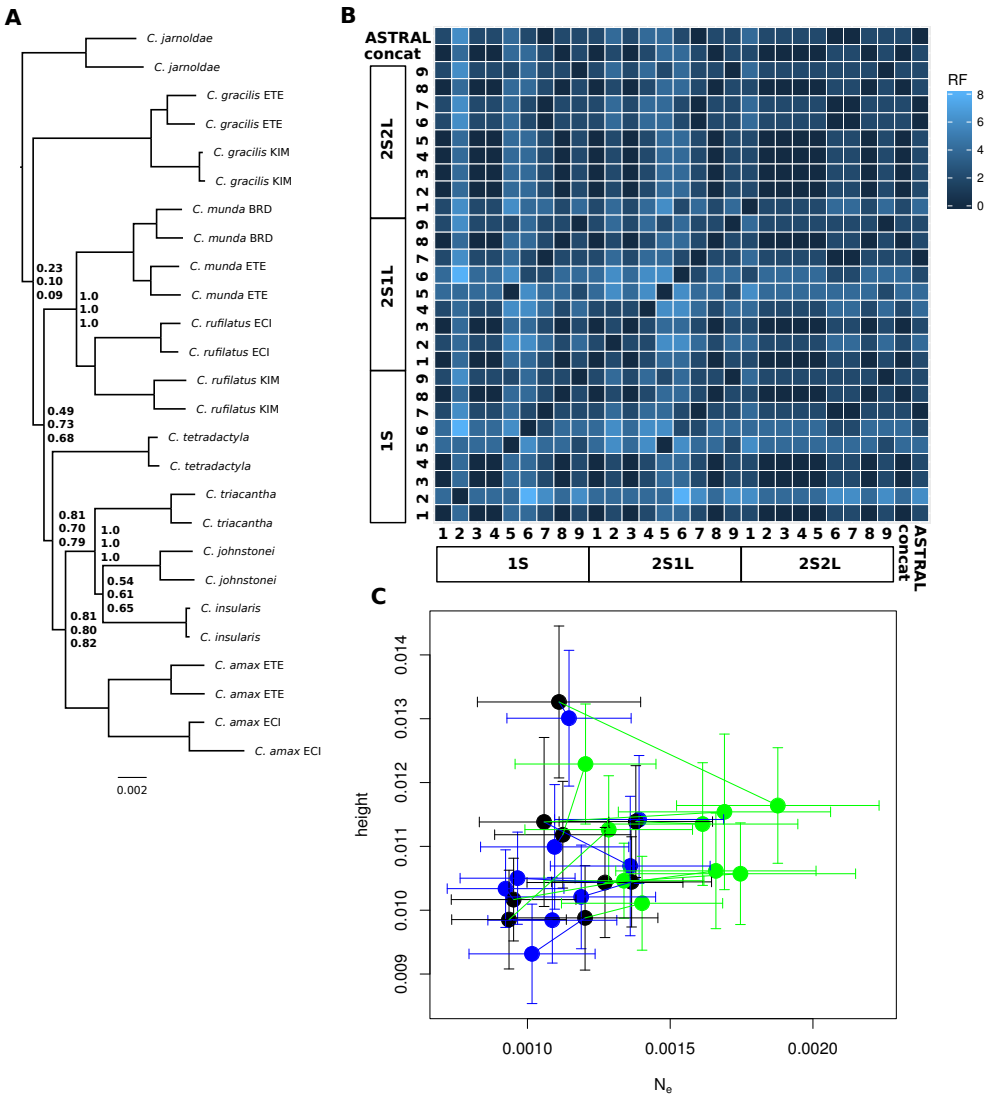


Figure 4.5 Outcomes of full MSC (StarBEAST2) analyses for a focal clade of 9 taxa of rainbow skinks, using three different 'treatments' for the assignment of alleles to each taxon. A. The phylogeny, as estimated using a concatenated alignment (also shown in Fig. 4.1), illustrates the level of divergence between lineages in the species (*C. amax*, *C. gracilis*, *C. munda* and *C. rufilatus*) whose sampling was manipulated in the experimental treatments (1S, 2S1L, 2S2L; described in text). Support values (on the nodes of A) show the support, across the 9 replicate posterior distributions, in the species trees estimated using the 1S, 2S1L, 2S2L treatments (respectively). B. The level of discordance (unrooted Robinson-Foulds distance) is shown among trees estimated with the 9 replicates of the tree treatments (maximum clade credibility trees), as well as the concatenation tree (A), and the topology estimated with ASTRAL (and shown in Fig. 4.2). C. The tree height and N_e values estimated by each of the replicates, shown as the mean (± s.d.) of 100 samples from the posterior distribution of each chain. Here treatment 1S shown in blue, 2S1L in black, and 2S2L in green.

Among genera, *Liburnascincus* and *Lygisaurus* consistently formed a clade, to the exclusion of *Carlia* (Fig. 4.1, 4.2). This finding is inconsistent with previous estimates of the relationship among these genera, which placed *Carlia* and *Lygisaurus* in a clade, to the exclusion of *Liburnascincus* [26,63], albeit with low support. Our phylogenomic data now provide strong support for generic relationships.

Within each genus, the relationships we inferred among species were often concordant with previous studies, but with notable exceptions. For the genus *Carlia*, our analyses support or modify several monophyletic groups identified by Dolman and Hugall [26], which then informed a set of 'species-groups' that were nominated by Zug [19], and extended by Hoskin [64]. For instance, we observe a clade from north Queensland corresponding to the *fusca* species group [19], containing a New Guinea representative of the *fusca* group (*C. waro* sp.), *C. sexdentata*, *C. longipes* and *C. quinquecarinata*. However, in contrast to previous studies, we find this group is part of a larger clade that includes *C. storri*, *C. schmeltzii* and *C. rostralis* (Fig. 4.1). In particular, the placement of *C. rostralis* in this clade deviates substantially from previous observations [26] and assignments [19].

The northeast Queensland rainforest-dwelling *rhomboidalis* species group [19] was supported as a clade in our analysis, consisting of *C. rhomboidalis* and *C. rubrigularis*. We also estimated a tree that places newly described *C. wundalthini* in this clade, as suggested by Hoskin [64]. We find that *C. wundalthini* is sister to the northern lineage of *C. rubrigularis*, while the southern lineage of *C. rubrigularis* is sister to *C. rhomboidalis*, confirming initial multilocus analysis by Dolman and Moritz [27]. A taxonomic revision of this group is in progress (S. Singhal et al., unpublished).

We inferred a widespread eastern Australian clade within *Carlia* containing the species *C. vivax*, *C. dogare*, *C. pectoralis*, *C. decora*, *C. rubigo* and *C. inconnexa*. Three of these species, *C. decora*, *C. rubigo* and *C. inconnexa*, were formerly included in *C. pectoralis* [34]. *Carlia rubigo* and *C. inconnexa* formed a clade with C. pectoralis, which supports morphological data suggesting these are three sister species [34]. The two *C. pectoralis* samples in the concatenated tree (Fig. 4.1) are not monophyletic; however, both '*C. pectoralis'* samples come from an area of uncertainty between the known distributions of C. *pectoralis* and *C. rubigo* [34], and it is possible that sample ABTC76957 (Blackdown Tableland) is a misidentified C. rubigo. *Carlia* decora was sister to *C. vivax* in our phylogenies, supporting previously noted morphological similarity and species group assignment [34,64].

Finally, for *Carlia*, we found substantial support for a clade containing the remaining nine taxa: eight of these are from the monsoonal tropics of northern Australia, but one, *C. tetradactyla*, is widespread in the southeast of the continent. Different methods sometimes produced conflicting estimates for relationships within this clade, particularly near its base. For instance, concatenation and summary coalescent (ASTRAL) analyses inferred strongly that *C. jarnoldae* is the sister to all other members of this clade (Figs 4.1, 4.2), whereas StarBEAST2 analyses of this focal clade (Fig. 4.5a), which often placed *C. gracilis* sister to the rest of the clade, or supported a clade consisting of *C. jarnoldae* and *C. gracilis*, albeit with

low posterior probabilities for these clades. Other relationships were strongly supported across all analyses. We found strong and consistent support for *C. munda* and *C. rufilatus* as sister taxa, contradicting previous observations [26]. A clade containing *C. insularis*, *C. johnstonei*, and *C. triacantha* / *C. isostriacantha* was strongly supported across all analyses. Within this clade, we inferred different topologies. For instance, the concatenation analysis shown in Fig. 4.1 inferred that *C. johnstonei* and *C. insularis* are sister taxa (consistent with [21]), while the MSC analysis shown in Fig. 4.2 inferred that *C. johnstonei* and *C. triacantha* / *C. isostriacantha* are sister taxa (again with equivocal node support). Here the key point, supported by both analyses, is that the two divergence events separating these three lineages occurred in relatively quick succession.

Within *Lygisaurus*, we inferred two major clades, the smaller of which consisted of *Ly. sesbrauna*, samples allied to *Ly. macfarlani* from northern Australia, and samples allied to *Ly. curtus* from Papua New Guinea. The larger clade contained all the other sampled taxa of *Lygisaurus*, as well as '*Carlia rimula*' (as noted above). This is interesting because *Lygisaurus* therefore consists of a widespread and diverse clade of north-east Australian species, occupying a variety of habitats from rainforest to arid areas and rock outcrops, and a clade that occurs in New Guinea and far north-eastern Australia. The two clades overlap in mesic forest of Cape York Peninsula.

In *Liburnascincus*, the widespread *Li. mundivensis* is most divergent, excluded from a clade containing the narrowly distributed species *Li. scirtetis, Li. coensis* and *Li. artemis* (Fig. 4.1, 4. 2), though the relationships within the latter clade were not well resolved in our analyses (Fig. 4.2). *Liburnascincus mundivensis* is found in rocky habitat over a large area of north-east Australia, whereas the other three species have highly localized distributions around boulder-fields and isolated rocky ranges on Cape York. Morphological and mtDNA similarity had previously been noted between *Li. artemis* and *Li. mundavensis* [65] but our phylogenies clearly place *Li. artemis* in a clade with the other highly localised species.

4.5 Conclusions

This study provides a well-resolved estimate of the phylogeny of the Australian rainbow skinks. Diversification processes in this group are of great interest given their habitat and phenotypic diversity in tropical Australia, often involving many sympatric species [24]. The tree is based on data from a large number of loci, and is largely robust to different inferential approaches. It provides strong support for many previously identified clades and resolved others that have remained uncertain despite previous morphological and genetic studies. This provides the foundation for resolving the evolution and taxonomy of this group. Here we have advanced the taxonomy by formally reassigning one species (*Carlia rimula* to *Lygisaurus rimula*).

Our phylogeny also highlights the extent of deep lineage divergence within some species, often with little corresponding morphological differentiation, and this will be followed by detailed taxonomic studies. Here we explored how this intraspecific diversity might affect

phylogenetic estimation, by performing an experiment where samples from diverged intraspecific lineages were deliberately placed together in 'taxa' for MSC analyses, violating the assumption of panmixia. We did not observe a substantial effect of this experimental treatment on the topology of inferred trees, and suggest that this might be an interesting topic for a future simulation study.

Finally, our phylogeny will provide an enhanced context for studies examining trait evolution of the rainbow skinks, their adaptation to varied habitats, and their biogeographic patterns. Additionally, we expect future studies will improve the resolution of clade memberships of *Carlia* and *Lygisaurus* from New Guinea and Wallacea, providing further insights into diversification processes across the region.

4.6 References

[1] Schluter D. The ecology of adaptive radiation. Oxford, Oxford UK; 2000.
[2] Morlon H. Phylogenetic approaches for studying diversification. Ecol. Lett. 2014; 17: 508-525.
[3] Maddison WP. Gene trees in species trees. Sys. Biol. 1997; 46: 523-536.
[4] Pamilo P, Nei M. Relationships between gene trees and species trees. Mol. Biol. Evol. 1988; 5: 568-583.
[5] Liu L, Pearl DK, Brumfield RT, Edwards SV. Estimating species trees using multiple-allele sequence data. Evolution. 2008; 62: 2080-2091.
[6] Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. Mol Biol Evol. 2010; 27: 570-580.
[7] Ogilvie HA, Heled J, Xie D, Drummond AJ. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Sys Biol. 2016; 65: 381-396.
[8] Ogilvie HA, Bouckaert RR, Drummond AJ. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. 2017. Mol. Biol. Evol.; 34: 2101-2114.
[9] Liu L, Wu S, Yu L. Coalescent methods for estimating species trees from phylogenomic data. J Sys Evol. 2015; 53: 380-390.
[10] Mirarab S, Bayzid MS, Warnow T. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Sys Biol. 2016; 65: 366-380.
[11] Mirarab S, Warnow T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 2015; 31: i44–i52.
[12] Kubatko LS Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Sys. Bio. 2007; 56: 17-24.
[13] Rabosky DL, Donnellan SC, Talaba AL, Lovette IJ. Exceptional among-lineage variation in diversification rate during the radiation of Australia's most diverse vertebrate clade. Proc R Soc B. 2007; 274: 2915-2923.
[14] Skinner A, Hugall AF, Hutchinson MN. Lygosomine phylogeny and the origins of Australian scincid lizards. J Biogeog. 2011; 38: 1044-1058.
[15] Brandley MC, Bragg JG, Singhal S, Chapple DG, Jennings CK, Lemmon AR, Lemmon EM, Thompson MB, Moritz C. Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian Eugongylus group scincid lizards. BMC Evol. Biol. 2015; 15:62.
[16] Blom MPK, Bragg JG, Potter S, Moritz C. Accounting for uncertainty in gene tree estimation: summary coalescent species tree inference in a challenging radiation of Australian lizards. Sys Biol. 2016; 66:352-366.
[17] Ingram G, Covacevich JA. Revision of the genus *Lygisaurus* de Vis (Scincidae: Reptilia) in Australia. Memoirs of the Queensland Museum. 1988; 25: 335–354

[18] Ingram G, Covacevich JA. Revision of the genus *Carlia* (Reptilia, Scincidae) in Australia and comments on *Carlia bicarinata* of New Guinea. Memoirs of the Queensland Museum. 1989; 27: 443–490

[19] Zug GR. An outlying *Carlia* population from Java and comments on species groups within the genus *Carlia* (Reptilia: Squamata: Scincidae). Proceedings of the California Academy of Sciences 2010; 61: 389-408.

[20] Hoskin CJ and Couper PJ. A new skink (Scincidae: *Liburnascincus*) from rocky habitat on Cape York, northeast Australia. Zootaxa 2015; 3994: 222-234.

[21] Afonso Silva AC, Santos N, Ogilvie HA, Moritz C. Validation and description of two new northwestern Australian Rainbow skinks with multispecies coalescent methods and morphology. PeerJ. 2017; 5: e3724

[22] Goodman BA. Divergent morphologies, performance, and escape behaviour in two tropical rock-using lizards (Reptilia: Scincidae). Biol J Linn Soc. 2007; 91: 85-98.

[23] Goodman BA, Miles DB, Schwazkopf L. Life on the rocks: habitat use drives morphological and performance evolution in lizards. Ecology 2008; 89: 3462-3471.

[24] Dolman G, Stuart-Fox D. Processes driving male breeding colour and ecomorphological diversification in rainbow skinks: a phylogenetic comparative test. Evol. Ecol. 2010; 24: 97-113.

[25] Stuart-Fox DM, Hugall AF, Moritz C. A molecular phylogeny of rainbow skinks (Scincidae: *Carlia*): taxonomic and biogeographic implications. Aust J Zool. 2002; 50: 39-51.

[26] Dolman G, Hugall AF. Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). Mol Phy Evol. 2008; 49: 782-794.

[27] Dolman G, Moritz C. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). Evolution 2006; 60: 573-582.

[28] Singhal S, Moritz C. Testing hypotheses for genealogical discordance in a rainforest lizard. Mol Ecol. 2012; 21: 5059-5072.

[29] Potter S, Bragg JG, Peter BM, Bi K, Moritz C. Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. Molecular Ecology 2016; 25: 1367-1380.

[30] Afonso Silva AC, Bragg JG, Potter S, Fernandes C, Coelho MM, Moritz C. Tropical specialist versus climate generalist: diversification and demographic history of sister species of *Carlia* skinks from northwestern Australia. Molecular Ecology. 2017. 26; 4045-4058.

[31] Potter S, Xue AT, Bragg JG, Rosauer DF, Roycroft EJ, Moritz C. Pleistocene climatic changes drive diversification across a tropical savannah. Molecular Ecology. In Press. doi: 10.1111/mec.14441

[32] Singhal S, Moritz C. Reproductive isolation between phylogeographic lineages scales with divergence. Proc R Soc B. 2013; 280: 20132246.

[33] Donnellan SC, Couper PJ, Saint KM, Wheaton L. Systematics of the *Carlia* '*fusca*' complex (Reptilia: Scincidae) from northern Australia. Zootaxa 2009: 1-31.

[34] Hoskin CJ, Couper PJ. Description of two new *Carlia* species (Reptilia: Scincidae) from north-east Australia, elevation of *Carlia pectoralis inconnexa* Ingram and Covacevich 1989 to full species status, and redescription of *Carlia pectoralis* (de Vis 1994). Zootaxa 2012; 3546: 1-28.

[35] Gnirke A, Melnikov A, Maguire J et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotech. 2009; 27: 182-189.

[36] Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 2012; 13: 403.

[37] Bragg JG, Potter S, Bi K, Moritz C. Exon capture phylogenomics: efficacy across scales of divergence. Mol Ecol Res. 2016; 16: 1059-1068.

[38] Sunnucks P, Hales DF. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Mol Biol Evol. 1996; 13: 510-524.

[39] Myer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010, doi:10.1101/pdb.prot5448
[40] Trifonov VA, Vorobieva NN, Rens W. FISH with and without COT1 DNA. In T Liehr, editor. Fluorescence In Situ Hybridization (FISH): Application Guide. Springer, Berlin. 2009. pp. 99-109.
[41] Singhal S. De novo transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. Mol Ecol Res. 2013; 13: 403-416.
[42] Altshul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403-401.
[43] Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 2008; 18: 821-829.
[44] Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Research 1999; 9: 868-877.
[45] Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 2005; 6: 31.
[46] Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012; 9: 357–359.
[47] McKenna AH, Hanna M, Banks E et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 2010; 20: 1297-1303.
[48] Ranwez V, Harispe S, Delsuc F, Douzery EJ. MACSE: Multiple alignment of coding sequences accounting for frameshifts and stop codons. PLoS ONE 2011; 6: e22594.
[49] Capella-Gutierrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009; 25: 1972-1973.
[50] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014; 30: 1312-1313.
[51] Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 2013; 497: 327-331.
[52] Manthey JD, Campillo LC, Burns KJ, Moyle RG. Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in Cardinalid tanagers (Aves, Genus: *Piranga*). Sys Biol. 2016; 65: 640-650.
[53] Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL. Analysis of a rapid evolutionary radiation using ultraconserved elements (UCEs): evidence for a bias in some multispecies coalescent methods. Sys Biol. 2016; 65: 612-627.
[54] Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol. 2015; 32: 268-274.
[55] Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013; 30: 1188-1195.
[56] Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol Biol. 2014; 14: 82.
[57] Jarvis ED, Mirarab S, Aberer AJ et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 2014; 346: 1320-1331.
[58] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2016. Vienna, Austria.
[59] Schliep KP. Phangorn: phylogenetic analysis in R. Bioinformatics. 2011. 27; 592-593.
[60] Paradis E., Claude J., Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 2004; 20: 289-290.
[61] Solis-Lemus C, Yang M, Ane C. Inconsistency of species tree methods under gene flow. Syst. Biol. 2016. 65; 843-851.
[62] Ingram GJ, Covacevich J. Two new Lygosomine skinks endemic to Cape York Peninsula. In NC Stevens, A Bailey, editors. Contemporary Cape York, Royal Society of Queensland, Brisbane. 1980. pp. 45-48.

[63] Pyron RA, Burbrink FT, Wiens JJ. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol Biol. 2013; 13: 93.
[64] Hoskin CJ. A new skink (Scincidae: *Carlia*) from the rainforest uplands of Cape Melville, north-east Australia. Zootaxa 2014; 3869: 224-236.
[65] Hoskin CJ and Couper PJ. A new skink (Scincidae: *Liburnascincus*) from rocky habitat on Cape York, northeast Australia. Zootaxa 2015; 3994: 222-234.

Abbreviations
ILS: incomplete lineage sorting
MSC: multispecies coalescent
CL: complete loci (i.e., loci with alignments containing few missing data)
CIL: complete and informative loci (i.e., loci with alignments containing few missing data, and that are phylogenetically informative)
1S: experimental treatment – alleles from one sample
2S1L: experimental treatment – alleles from 2 samples from 1 lineage
2S2L: experimental treatment – alleles from 2 samples from 2 lineages

Declarations
Ethics approval and consent: This work was carried out under animal ethics approval through the Australian National University, protocol numbers A2012/14 and A2013/08.

Availability of data and materials: Raw sequence reads are available on the NCBI short read archive (SRA), in BioProject PRJNA289283. Processed data, including alignments, and code used for bioinformatic analyses are available in dryad repository: doi.org/10.5061/dryad.v1d32.

Funding: Funding for this research was provided by the Australian Research Council. The Australian research council did not have a role in the design of the study, the collection, analysis or interpretation of the data, or in writing the manuscript.

Authors' Contributions: SP, ACAS, CJH, and CM performed field work and facilitated access to required samples. SP and ACAS performed laboratory work. JGB and BYHB wrote portions of the analytical workflow. JBG performed the inferential analyses. All authors contributed to and approved the manuscript.

Acknowledgements: We thank Sonal Singhal, Ke Bi, Lydia Smith, and members of the Moritz Lab for discussion. We also thank Eric Rittmeyer for advice and access to samples, and Emily Roycroft for mitochondrial sequencing of two species. We are grateful to Huw Ogilvie for early access to StarBEAST2, and advice on its implementation. We thank the South Australian Museum, the Museum and Art Gallery of the Northern Territory, the Western Australian Museum, the Queensland Museum, and the Australian Museum for samples.

Supplementary Material

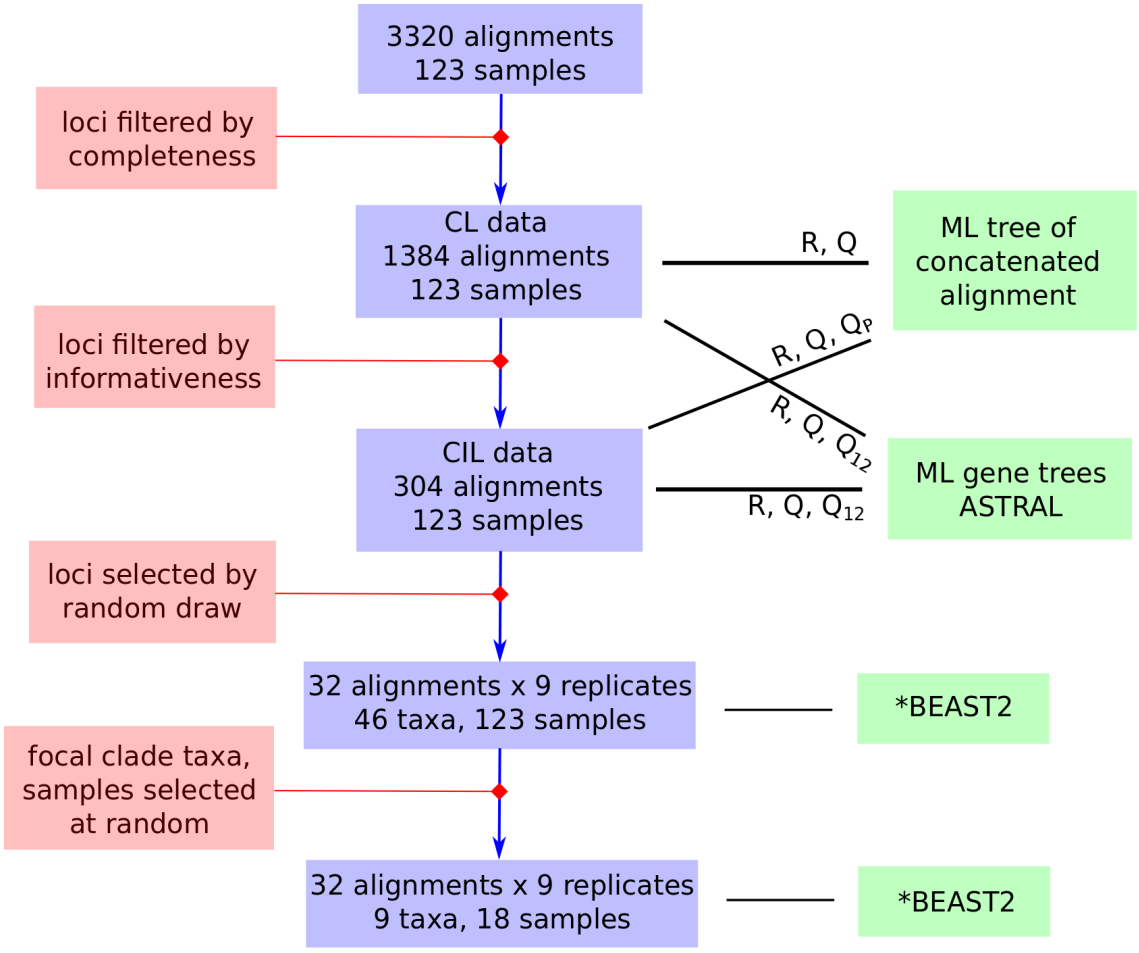


Fig. S4.1 A guide to the phylogenetic analyses performed in this study (see text for more details of methods). We sequenced a total of 3320 loci. After making alignments, we had 1384 loci with highly complete data, and designate these CL. Of these, we identified a set of 304 loci that were also highly phylogenetically informative, and these are designated CIL. We performed both concatenated and summary coalescent (ASTRAL) analyses of the CL and CIL data. We checked these analyses were robust to different analytical approaches by using different implementations (RAxML, R; IQTREE, Q), partitioning (Q_P) and excluding third codon positions (Q_{12}). We also chose 9 sets of 32 loci at random (without replacement) from the CL data, and for each of these, performed a species tree analysis using StarBEAST2. Finally, we performed an experiment that systematically manipulated the way alleles were selected to represent taxa in StarBEAST2 analyses, and ran these analyses for a focal clade of 9 taxa.

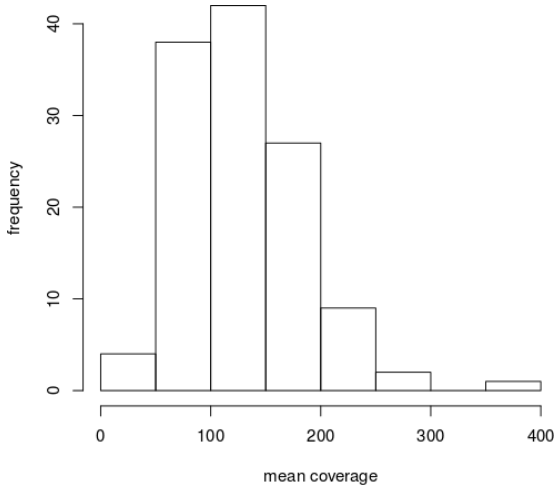


Fig. S4.2 Histogram showing the distribution of sequencing coverage (averaged across the assembled loci) for 123 skink samples.

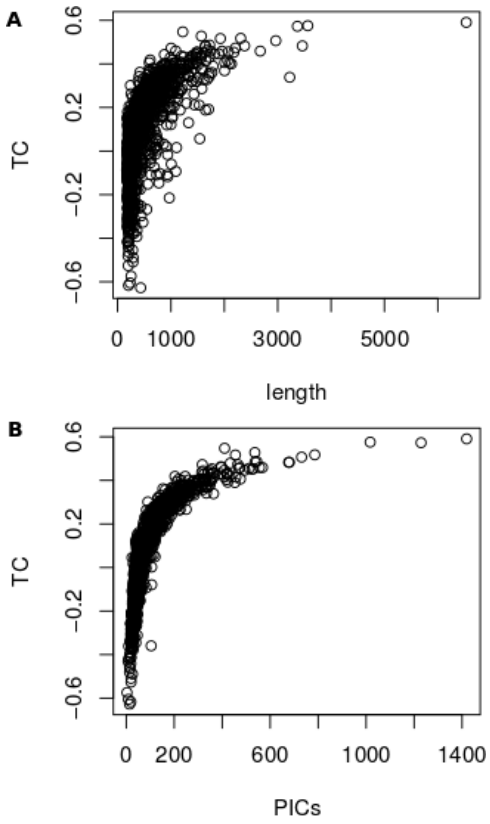


Fig. S4.3 The relationship between the relative tree certainty score (TC; including all conflicting bipartitions [50]) of the CL alignments, and their (A) length and (B) number of parsimony informative characters (PICs).

Table S4.1 Information about samples used in this phylogenomic study of the rainbow skinks. For each sample, the genus and species is given along with a tissue number and the collection where the sample was obtained. A short read archive BioSample number is also provided. Where samples have been used in published articles previously, a citation is given. The library name of the sample (used for sequencing) is provided. The column Species also corresponds to the 'taxon' to which samples were assigned in MSC analyses. For samples that were used in the 'Focal Clade' experiment, names of corresponding lineages are provided (in the Focal Clade column). NTM – Northern Territory Museum; SAM – South Australia Museum; WAM – Western Australia Museum; ANU – Australian National University; AM – Australian Museum; QM – Queensland Museum; JCU – James Cook University. * In StarBEAST2 analyses of the rainbow skink clade, these samples were included in the 'taxon' *Carlia triacantha*, prior to the description of *Carlia isostriacantha*.

Genus	Species	Tissue	Collection	BioSample (SRA)	Citation	Library	Focal Clade
Carlia	*amax*	ABTC29892	NTM	SAMN04420550	[29]	SP03_25	ECI
Carlia	*amax*	ABTC11961	SAM	SAMN04420538	[29]	SP03_26	
Carlia	*amax*	ABTC72760	SAM	SAMN04420560	[29]	SP03_30	
Carlia	*amax*	R36590	NTM	SAMN04420576	[29]	SP03_36	
Carlia	*amax*	ABTC29886	NTM	SAMN04420549	[29]	SP04_11	
Carlia	*amax*	ABTC29931	NTM	SAMN04420551	[29]	SP04_17	ECI
Carlia	*amax*	ABTC28693	NTM	SAMN04420544	[29]	SP04_18	
Carlia	*amax*	R114401	WAM	SAMN04420563	[29]	SP04_33	
Carlia	*amax*	R171890	WAM	SAMN04420569	[29]	SP04_34	
Carlia	*amax*	R28957	NTM	SAMN04420571	[29]	SP04_37	
Carlia	*amax*	A002942	QM	SAMN04420533	[29]	SP04_55	
Carlia	*amax*	ABTC72498	SAM	SAMN05207399	[31]	SP11_29	ETE
Carlia	*amax*	CCM1996	ANU	SAMN05207400	[31]	SP11_30	ETE
Carlia	*insularis*	R117953	WAM	SAMN06927852	[30]	AS01_45	INS
Carlia	*insularis*	R117967	WAM	SAMN06927853	[30]	AS01_46	INS
Carlia	*decora*	conx5115	QM	SAMN05207333		SP07_4	
Carlia	*decora*	conx5116	QM	SAMN05207353		SP08_4	
Carlia	*dogare*	ABTC32199	QM	SAMN05207334		SP07_5	
Carlia	*dogare*	ABTC32200	QM	SAMN05207354		SP08_5	
Carlia	*gracilis*	CCM0457	ANU	SAMN05207314	[31]	SP05_14	
Carlia	*gracilis*	ABTC30644	NTM	SAMN04420554	[29]	SP05_2	ETE
Carlia	*gracilis*	CCM1135	ANU	SAMN05207316	[31]	SP05_23	KIM
Carlia	*gracilis*	CCM0957	ANU	SAMN05207317	[31]	SP05_9	KIM
Carlia	*gracilis*	ABTC11793	SAM	SAMN05207392	[31]	SP11_1	ETE
Carlia	*gracilis*	ABTC28428	NTM	SAMN05207393	[31]	SP11_10	
Carlia	*gracilis*	CCM0958	ANU	SAMN05207406	[31]	SP11_6	
Carlia	*inconnexa*	J89138	QM	SAMN05207355		SP08_7	
Carlia	*inconnexa*	J89132	QM	SAMN05207391		SP10_56	
Carlia	*jarnoldae*	ABTC1107	QM	SAMN05207336		SP07_8	JAR
Carlia	*jarnoldae*	ABTC77191	SAM	SAMN05207356	[30]	SP08_8	JAR
Carlia	*johnstonei*	R171237	WAM	SAMN06927836	[30]	AS01_29	
Carlia	*johnstonei*	CCM0934	ANU	SAMN06927837	[30]	AS01_30	JOH
Carlia	*johnstonei*	R117726	WAM	SAMN06927838	[30]	AS01_31	JOH
Carlia	*johnstonei*	MCZA28608	WAM	SAMN06927854	[30]	AS01_47	
Carlia	*longipes*	ABTC11002	AM	SAMN05207337		SP07_9	
Carlia	*longipes*	ABTC51118	SAM	SAMN05207357		SP08_9	
Carlia	*munda*	R131750	WAM	SAMN05207312	[31]	SP04_43	
Carlia	*munda*	ABTC10845	AM	SAMN05207318	[31]	SP07_10	BRD
Carlia	*munda*	ABTC10846	AM	SAMN05207338	[31]	SP08_10	BRD
Carlia	*munda*	ABTC102927	QM	SAMN05207401	[31]	SP11_33	
Carlia	*munda*	ABTC72508	NTM	SAMN05207402	[31]	SP11_44	
Carlia	*munda*	ABTC28277	NTM	SAMN05207403	[31]	SP11_45	
Carlia	*munda*	ABTC10977	AM	SAMN05207404	[31]	SP11_46	ETE
Carlia	*munda*	ABTC28960	NTM	SAMN05207405	[31]	SP11_47	ETE
Carlia	*pectoralis*	ABTC76882	SAM	SAMN05207322		SP07_14	
Carlia	*pectoralis*	ABTC76957	SAM	SAMN05207342		SP08_14	
Carlia	*quinquecarinata*	ABTC102373	QM	SAMN05207323		SP07_15	
Carlia	*quinquecarinata*	ABTC102374	QM	SAMN05207343		SP08_15	
Carlia	*rhomboidalis*	ABTC80487	SAM	SAMN05207387		SP10_20	
Carlia	*rhomboidalis*	ABTC14188	SAM	SAMN05207388		SP10_28	
Carlia	*rostralis*	A006771	QM	SAMN05207372		SP09_26	

Genus	Species	Tissue	Collection	BioSample (SRA)	Citation	Library	Focal Clade
Carlia	*rostralis*	A006581	QM	SAMN05207373		SP09_27	
Carlia	*rubigo*	J89141	QM	SAMN05207325		SP07_17	
Carlia	*rubigo*	J89142	QM	SAMN05207345		SP08_17	
Carlia	*rubrigularis North*	SS33	ANU	SAMN03787456	[29]	SP03_6	
Carlia	*rubrigularis North*	ABTC11016	AM	SAMN05207324		SP07_16	
Carlia	*rubrigularis North*	ABTC11021	AM	SAMN05207344		SP08_16	
Carlia	*rubrigularis South*	SS46	ANU	SAMN03787434	[37]	SP02A_7	
Carlia	*rubrigularis South*	ABTC77086	SAM	SAMN05207386		SP10_18	
Carlia	*rufilatus*	CMWA35	ANU	SAMN05207313	[31]	SP04_9	KIM
Carlia	*rufilatus*	R27658	NTM	SAMN05207315	[31]	SP05_16	
Carlia	*rufilatus*	ABTC29321	NTM	SAMN05207394	[31]	SP11_16	ECI
Carlia	*rufilatus*	ABTC29295	NTM	SAMN05207395	[31]	SP11_17	ECI
Carlia	*rufilatus*	CCM0677	ANU	SAMN05207396	[31]	SP11_22	
Carlia	*rufilatus*	ABTC29941	NTM	SAMN05207397	[31]	SP11_24	
Carlia	*rufilatus*	ABTC28405	AM	SAMN05207398	[31]	SP11_28	KIM
Carlia	*schmeltzii*	ABTC11024	AM	SAMN05207326		SP07_18	
Carlia	*schmeltzii*	ABTC77118	SAM	SAMN05207346		SP08_18	
Carlia	*sexdentata*	ABTC10982	AM	SAMN05207327		SP07_19	
Carlia	*sexdentata*	ABTC10985	AM	SAMN05207347		SP08_19	
Carlia	*sp. (Waro)*	ABTC44734	AM	SAMN05207335		SP07_6	
Carlia	*storri*	A010492	QM	SAMN05207380		SP09_40	
Carlia	*storri*	ABTC53601	SAM	SAMN05207389		SP10_29	
Carlia	*tetradactyla*	ABTC11042	AM	SAMN05207328		SP07_20	TET
Carlia	*tetradactyla*	ABTC17099	SAM	SAMN05207348		SP08_20	TET
Carlia	*isostriacantha**	R168590	WAM	SAMN06927823	[21]	AS01_16	
Carlia	*triacantha*	R139010	WAM	SAMN06927841	[21]	AS01_34	TRI
Carlia	*triacantha*	ABTC29091	SAM	SAMN06927846	[21]	AS01_39	
Carlia	*triacantha*	CCM1859	NTM	SAMN06927848	[21]	AS01_41	TRI
Carlia	*isostriacantha**	ABTC29851	SAM	SAMN06927812	[21]	AS01_5	
Carlia	*vivax*	A006791	QM	SAMN05207376		SP09_3	
Carlia	*vivax*	A007267	QM	SAMN05207379		SP09_4	
Carlia	*wundalthini*	conx5328	JCU	SAMN05207374		SP09_28	
Carlia	*wundalthini*	conx5330	JCU	SAMN05207382		SP09_56	
Lampropholis	*coggeri*	SS60	ANU	SAMN03787435	[37]	SP02A_4	
Lampropholis	*coggeri*	SEW8451	ANU	SAMN03787429	[37]	SP02A_5	
Lampropholis	*guichenoti*	ABTC12335	SAM	SAMN05207329		SP07_28	
Lampropholis	*guichenoti*	ABTC79668	SAM	SAMN05207349		SP08_28	
Liburnascincus	*artemis*	conx5371	JCU	SAMN05207375		SP09_29	
Liburnascincus	*artemis*	conx5373	JCU	SAMN05207377		SP09_30	
Liburnascincus	*coensis*	A004566	QM	SAMN05207365		SP09_17	
Liburnascincus	*coensis*	A002088	QM	SAMN05207366		SP09_18	
Liburnascincus	*mundivensis*	ABTC10839	AM	SAMN05207319		SP07_11	
Liburnascincus	*mundivensis*	ABTC77119	SAM	SAMN05207339		SP08_11	
Liburnascincus	*scirtetis*	A002000	QM	SAMN05207367		SP09_19	
Liburnascincus	*scirtetis*	A004721	QM	SAMN05207369		SP09_20	
Lygisaurus	*aeratus*	ABTC10855	AM	SAMN05207381		SP09_5	
Lygisaurus	*aeratus*	A007261	QM	SAMN05207383		SP09_6	
Lygisaurus	*cf. curtus*	ABTC44460	AM	SAMN05207320		SP07_12	
Lygisaurus	*cf. curtus*	ABTC46164	AM	SAMN05207340		SP08_12	
Lygisaurus	*cf. macfarlani*	ABTC30000	NTM	SAMN05207331		SP07_30	
Lygisaurus	*cf. macfarlani*	ABTC29123	NTM	SAMN05207351		SP08_30	

Genus	Species	Tissue	Collection	BioSample (SRA)	Citation	Library	Focal Clade
Lygisaurus	*macfarlani*	conx5614	JCU	SAMN05207378		SP09_39	
Lygisaurus	*foliorum*	ABTC10971	AM	SAMN05207330		SP07_29	
Lygisaurus	*foliorum*	ABTC72910	SAM	SAMN05207350		SP08_29	
Lygisaurus	*laevis*	SEW8024	ANU	SAMN03787451	[37]	SP03_17	
Lygisaurus	*laevis*	A001848	QM	SAMN05207359		SP09_10	
Lygisaurus	*laevis*	A000355	QM	SAMN05207360		SP09_11	
Lygisaurus	*malleolus*	A006602	QM	SAMN05207361		SP09_12	
Lygisaurus	*malleolus*	A006770	QM	SAMN05207362		SP09_13	
Lygisaurus	*parrhasius*	ABTC31977	QM	SAMN05207321		SP07_13	
Lygisaurus	*parrhasius*	ABTC31978	QM	SAMN05207341		SP08_13	
Carlia	*rimula*	A004565	QM	SAMN05207358		SP09_1	
Carlia	*rimula*	A004595	QM	SAMN05207368		SP09_2	
Lygisaurus	*rococo*	LR7	JCU	SAMN05207332		SP07_31	
Lygisaurus	*rococo*	LR11	JCU	SAMN05207352		SP08_31	
Lygisaurus	*sesbrauna*	A004711	QM	SAMN05207363		SP09_15	
Lygisaurus	*tanneri*	A004762	QM	SAMN05207364		SP09_16	
Lygisaurus	*tanneri*	ABTC32197	QM	SAMN05207390		SP10_30	
Lygisaurus	*zuma*	A000129	QM	SAMN05207384		SP09_8	
Lygisaurus	*zuma*	A007790	QM	SAMN05207385		SP09_9	
Pygmaeascincus	*timlowi*	A001585	QM	SAMN05207370		SP09_21	
Pygmaeascincus	*timlowi*	A007378	QM	SAMN05207371		SP09_22	

Chapter V

DISSECTING INTROGRESSION USING EXON CAPTURE DATA FROM SYMPATRIC RAINBOW SKINKS ACROSS THE AUSTRALIAN MONSOONAL TROPICS

Ana C. Afonso Silva[1,2], Sally Potter[1], Mozes P. K. Blom[3], Jason G. Bragg[4], Carlos Fernandes[2], Maria Manuela Coelho[2] and Craig Moritz[1]

[1]Division of Ecology and Evolution, Research School of Biology and Centre for Biodiversity Analysis, Australian National University, Acton, ACT, Australia.
[2]cE3c - Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.
[3]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden
[4]Herbarium of NSW, Royal Botanic Gardens & Domain Trust, Sydney

The original work described in this chapter is in preparation to be submitted in the journal Molecular Ecology.

Abstract

The increasing availability of sequence data in the age of genomics has provided stronger evidence of introgression between diverging species than previously thought. Additionally, new statistical tools have been developed to improve our ability to distinguish incomplete lineage sorting from events of past introgression, allowing us to understand how introgression has influenced species evolutionary histories and the role of hybridization on species persistence in more or less unstable environments.

Here we study introgression patterns between six closely related and often sympatric rainbow skink species from the Australian Monsoonal Tropics, a region with high biodiversity and prevalence of morphologically cryptic lineages within species. Using exon capture sequence data from >1700 loci, we test for recent admixture, ancient introgression between species and among lineages that overlap in regions with contrasting climatic histories.

Despite broad overlap in geographic distributions, we detected a negligible signature of recent admixture between species despite only moderate genetic divergence, suggesting strong reproductive isolation among these species. However, we did find support for several ancestral introgression events, likely early in the divergence histories of these taxa. The few instances of introgression detected in the comparisons of regions support the potential role of hybridization in a region of higher climatic instability.

This study demonstrates that introgression might have important consequences in species evolutionary histories, through the speciation process to species' responses to environmental instability.

Keywords: Introgression, Australian Monsoonal Tropics, rainbow skinks, phylogenetic networks, ABBA-BABA

5.1 Introduction

As we investigate evolutionary histories within and among species with increasingly available and affordable multilocus and genomic data, a major challenge arises in the presence of conflicting information from different gene trees (Degnan & Rosenberg 2009; Edwards et al. 2016). The two main causes of gene tree discordance in eukaryotes are incomplete lineage sorting (ILS) and introgression. ILS occurs when the most-recent common ancestor for a genealogy exists before speciation begins, allowing the genealogical relationship of a gene or genome region to differ from the underlying species branching pattern. Introgression is the process in which genetic material is transferred from one population or species to another by hybridization and occurs when introgressed variants persist in the new host through drift or selection (Shapiro et al. 2016). Reticulate evolution can be a consequence of substantial introgression, and this has long been recognized as important in plants (Anderson & Stebbins Jr. 1954; Rieseberg 1995). With the increase of genome scale analyses, historical introgression among species is now increasingly reported in animals as well (Hedrick 2013; Mallet et al. 2016). Currently widely described, reticulate evolution can be a source of novel gene combinations allowing adaptive evolution – e.g. in hominids (Green et al. 2010; Reich et al. 2011; Tung & Barreiro 2017), *Anopheles* mosquitoes (Fontaine et al. 2015), *Heliconius* butterflies (The *Heliconius* Genome Consortium 2012), geese (Ottenburghs et al. 2016), bears (Kumar et al. 2017), cichlids (Brawand et al. 2014) and *Xiphophorus* fishes (Cui et al. 2013).

Introgression has mostly been studied to understand divergence histories, adaptive divergence and the speciation process (Harrison & Larson 2014), either through hybrid zones studies and associated cline analyses (reviewed in Gompert et al. 2017), or by measuring the amount of gene flow between lineages or closely related species (Edwards et al. 2016). Commonly, recent admixture is detected using clustering methods like STRUCTURE (Pritchard et al. 2000) based on assignments of individual ancestry (additional current genomic approaches reviewed in Payseur & Rieseberg 2016). In contrast, methods with isolation-with-migration models, including approaches based on the site frequency spectrum, explore rates and direction of gene flow during divergence, and test for secondary contact of sister species (reviewed in Pinho & Hey, 2010; Sousa & Hey, 2013). Lately, there has also been a considerable development of methods that can detect reticulation at deeper phylogenetic scales - between non-sister taxa and their ancestors - while distinguishing ILS from introgression. These have been explored using either SNP-based approaches such as the ABBA-BABA test (Green et al. 2010; Durand et al. 2011; Eaton & Ree 2013; Pease & Hahn 2015; Ward & van Oosterhout 2016), or sequence-based approaches under the multispecies network coalescent (MSNC) model with maximum likelihood methods (Yu et al. 2014; Yu & Nakhleh 2015; Solís-Lemus & Ané 2016) or, more recently, fully Bayesian methods (Wen & Nakhleh 2017; Zhang et al. 2017).

5.1.1 Introgression and environmental instability

At small spatial and temporal scales, it has been shown that environmental change can increase the likelihood of hybridization between populations or closely related species (reviewed in Anderson 1948 and more recent examples Seehausen et al. 2008; Arnold & Martin 2010). This, depending on its magnitude, could have lasting effects due, for example, to an increase in genetic diversity of a population/species or the transfer of potentially beneficial alleles. Examples of species showing evidence of introgression that likely result from past hybridization associated with environmental change range from species that changed their distribution following the LGM, especially in the Northern Hemisphere (e.g. Iberian hare, Marques et al. 2017), to species flocks and adaptive radiations such as the African lake cichlids. Seehausen (2004), for instance, suggested that large-scale climatic and tectonic changes resulted in environmental fluctuations leading to habitat variability, facilitating hybridization that ultimately promoted cichlids speciation and rapid diversification events (see also e.g. Ivory et al. 2016; Meier et al. 2017). As tropical ectotherms typically have narrow physiological limits (Deutsch et al. 2008), they could be especially prone to range shifting through glacial cycles, forming novel communities if species respond idiosyncratically (Potter et al. 2017).

The study of currently sympatric and closely-related ectotherms can be helpful to understand the above processes and provides an opportunity to quantify recent and historical gene flow between species in relation to their genetic, phenotypic and ecological divergence. Genetic divergence among lineages can be an effective predictor of suppression of gene flow (Roux et al. 2016), including in species related to those studied here (Singhal & Bi 2017). Phenotypic divergence can also suppress introgression, especially where traits involved in mate recognition have become distinct (Nosil 2012). Ecological divergence can also suppress gene flow through habitat-associated assortative mating or selection against hybrids (Schluter 2000; Nosil 2012).

5.1.2 AMT Rainbow skinks

Following Edwards et al. (2016), we apply multiple approaches to test for contemporary or historical reticulation among the Australian Monsoonal Tropics (AMT) rainbow skinks, a group of six closely-related species (Bragg et al. 2018) with overlapping ranges across most of the AMT, especially in the Kimberley (northwest) and Top End (central north) regions of Australia. This clade of mid to late Miocene origin (Dolman & Hugall 2008) includes multiple leaf-litter species (*Carlia* amax, *C. gracilis*, *C. johnstonei*, *C. munda*, *C. rufilatus* and *C. triacantha*) with similarities in body size and life history, but vary somewhat in the breeding colors of males (James & Shine 1988; Dolman & Stuart-Fox 2010). These species further differ in observed climatic niche breadth and habitat preferences, with *C. gracilis* and *C. johnstonei* being more restricted to closed and mesic habitats; *C. rufilatus* and *C. amax* occurring in more intermediate habitats; whereas *C. munda* and *C. triacantha (s. l.)* extend across both mesic and arid habitats (Fig. 5.1, Dolman & Stuart-Fox, 2010; Wilson & Swan, 2013; Cogger,

2014). Phylogeographic diversity and demographic history within *Carlia* species have recently been studied using sequence data from mtDNA and >2000 exons, leading to the identification of several genetically divergent lineages within each species (Potter et al. 2016, 2017; Afonso Silva et al. 2017a), amounting to a total of 18 lineages across the six species (Fig. 5.1). Progress with taxonomic revision varies across taxa; in both *C. triacantha* and *C. johnstonei*, subsequent genetic and morphological analyses resulted in the formal recogni-

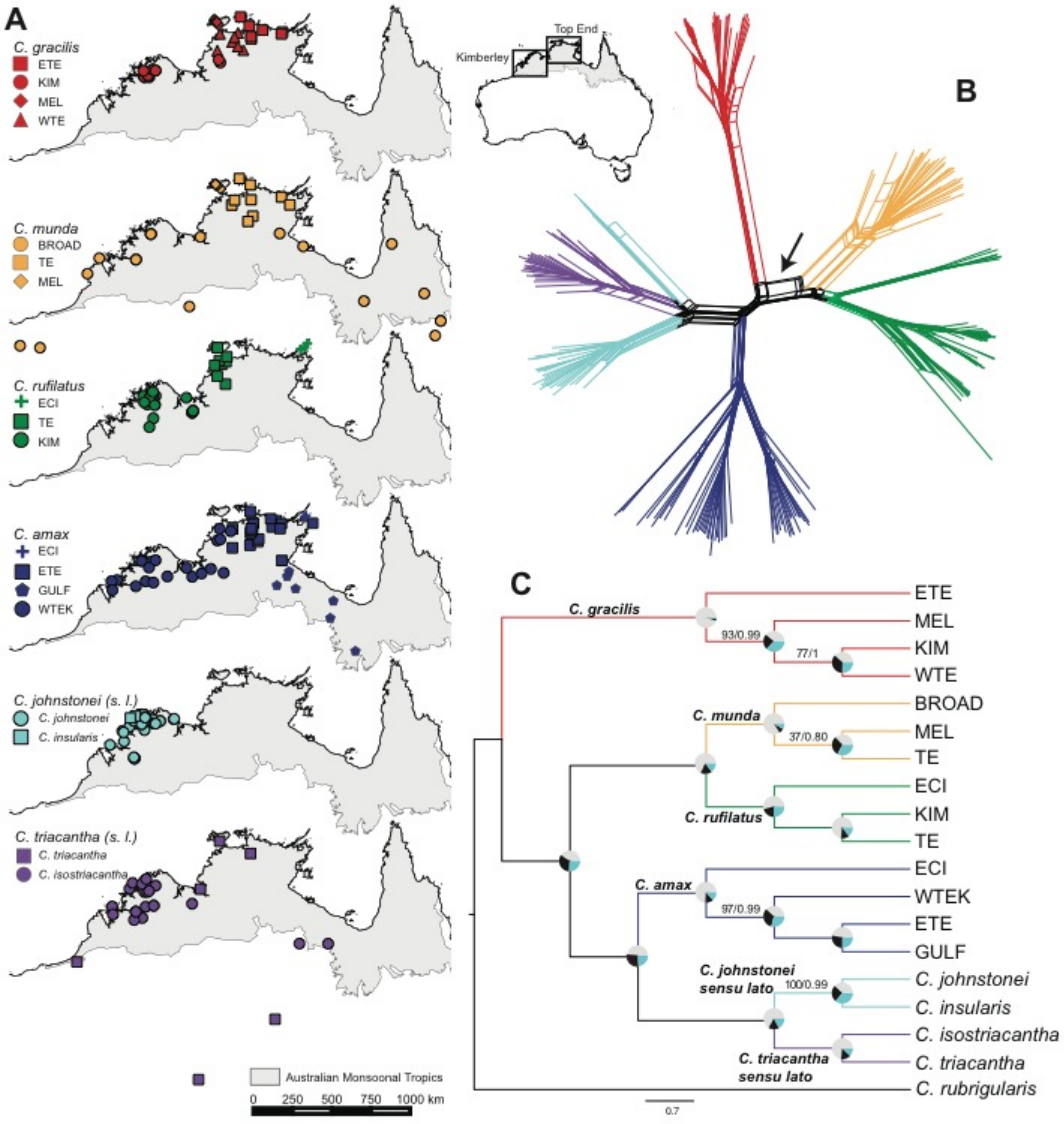


Figure 5.1 A. Geographic distribution of samples of the studied species B. SplitsTree Neighbor-Net network including all analysed individuals, which are coloured by species (colour codes as in A). C. ASTRAL species tree. Numbers on branches separated by a slash are, respectively, bootstrap percentages and local posterior probabilities based on ML gene trees. Branches without values were fully supported by both measures. Pie charts on nodes represent relative frequencies of the three quartet topologies around a branch. Species are represented by different colours: red – *C. gracilis*, yellow – *C. munda*, green – *C. rufilatus*, dark blue – *C. amax*, light blue - *C. johnstonei (s. l.)* and purple - *C. triacantha (s. l.)*. Each lineage is represented by a different symbol and lineage names are as in Potter et al. 2016: TE (Top End; ETE – eastern; WTE – western), KIM (Kimberley), MEL (Melville Island), BROAD (widespread across regions), ECI (English Company Islands), GULF (gulf of Carpentaria) and WTEK (western Top End and Kimberley).

tion of deeply divergent lineages as distinct species (*C. isostriacantha* and *C. insularis*, respectively; Afonso Silva et al. 2017b), but lineages with equal to or deeper divergence in the other taxa await further investigation. Therefore, to enable consistent treatment of taxa, we treat *C. isostriacantha* and *C. insularis* as divergent lineages within *C. triacantha* (*sensu lato*) and *C. johnstonei (s. l.)*, respectively.

Other than the Kimberley endemic *C. johnstonei (s. l.)*, each taxon has distinct lineages in the Top End vs Kimberley (sometimes extending to the western Top End). These regions seem to differ in climates now and through the last climate cycle, especially for precipitation, with the Kimberley being more arid and with even more restrictive conditions historically than the Top End; this associated with more disparate timings of population fluctuations in the Kimberley (Potter et al. 2017).

Past work in the genus *Carlia* has revealed gene flow between lineages of the same species. Potter et al. (2017) found, for most AMT *Carlia*, some evidence of admixture, additionally identifying a few hybrids between intraspecific lineages in contact zones and substantial levels of inferred migration between mainland lineages through their divergence history. While Afonso Silva et al. 2017a described the divergence of lineages within *C. johnstonei (s. l.)* and *C. triacantha (s. l.)* in the presence of gene flow, although with reduced migration rates. Additionally, and as a useful benchmark here, comparative genomic analyses of hybrid zones between intraspecific lineages of related species in the Wet Tropics, including two divergent lineages of *Carlia* rubrigularis, found that reproductive isolation increased with divergence (Singhal & Moritz 2013; Singhal & Bi 2017). It has not yet been determined whether there is also introgression among other species of *Carlia*, but this is entirely possible as genomic divergence among these sympatric AMT species is within the range where hybridization (with limited introgression) is observed between intraspecific lineages of Wet Tropics taxa (see Results). Further, a recent phylogenomic analysis for the genus revealed conflicting topologies across coalescent species-tree and concatenation methods (Bragg et al 2018), which may indicate effects of reticulation among non-sister taxa (Solís-Lemus et al. 2016).

Here we test for introgression between the sympatric species of AMT *Carlia* and in relation to regional differences in climatic history and, hence, potential range shifting. Specifically, we combine new and published sequence data for ~1700 exons with multiple new statistical methods for detecting introgression. We tested for i) recent admixture among species, and then for historical introgression while accounting for ILS at two different levels: ii) between species before intraspecific lineage divergence, and iii) between lineages from different species that broadly overlap. To do this, for each species we focused on a subset of individuals from the major lineages present in the Kimberley and in the Top End. We used this strategy to test for more introgression in the Kimberley, a region that has had a more dynamic climatic history than the more mesic and consistently suitable Top End (Potter et al. 2017). More specifically, we hypothesize that given the Pleistocene climatic instability in the Kimberley and the climatic niche breadth differences between species of rainbow skinks

(Potter et al. 2017; Afonso Silva et al. 2017a), there could have been more opportunities for hybridization events and substantial introgression in the Kimberley than in the more stable Top End. From the species-level comparisons, we also explored whether there is greater potential for introgression among more closely related species, as expected from comparative hybrid zone analyses of variously related taxa (Singhal & Moritz 2013; Singhal & Bi 2017).

5.2 Material and Methods

Samples were sequenced with an in-solution exon capture approach. For *C. amax* (A), *C. gracilis* (G), *C. munda* (M) and *C. rufilatus* (R) sequence data were from Potter et al. (2016, 2017), for *C. johnstonei* (J) and *C. triacantha* (T) (both *sensu lato*) data were from Afonso Silva et al. (2017a), and additional *Carlia* spp. samples from Bragg et al. (2016) were also included. These were supplemented with 72 new samples and representing all six species, for a total of 228 samples plus an outgroup (*C. rubrigularis*, based on previous work: Dolman & Hugall, 2008; Potter et al., 2017).

Total DNA was extracted using the salting-out method of Sunnucks & Hales (1996), and the genomic libraries for exon capture were constructed following the protocol of Meyer & Kircher (2010), using the modifications of Bi et al. (2012). The exon capture probe set was designed from transcriptomes of three related genera of skinks (*C. rubrigularis*, *Lampropholis coggeri*, *Saproscincus basiliscus;* Singhal, 2013) with exons selected from orthologs annotated against the *Anolis* genome. A total of 3320 target exon (> 200 bp) probes were designed and synthesized as a SeqCap EZ Developer library by Roche NimbleGen (see Bragg et al. 2016 for more information). For a single hybridization reaction, we pooled 56 individuals into a total of 1200 ng of barcoded genomic libraries, combined in equimolar ratios prior to hybridization with probes, index-specific blocking oligos and a specific skink Cot-1 DNA (see Bragg et al. 2016). After clean up, we ran two independent enrichment PCRs using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) for 17 cycles. We then performed a qPCR using specific primers (from Potter et al. 2016) to evaluate capture success, by confirming that loci targeted in the exon capture were enriched while non-targeted loci were de-enriched. To quantify and assess the quality of the hybridization, we measured pre- and post-capture library concentrations using a Bioanalyzer 2100 (Agilent Technologies, Inc.). After confirming the successful hybridization, the enriched libraries were sequenced on a single Illumina HiSeq 2500 lane (100 bp paired-end reads) at the ACRF Biomolecular Resource Facility, ANU. The new sequences were added to the same NCBI Short Read Archive (BioProject PRJNA289283) containing the samples used from previous studies (BioSample numbers in Appendix S5.1, Supplementary material).

5.2.1 Bioinformatics methods

The Illumina sequencing reads were processed following Bragg et al. (2016). Briefly, the workflow starts with a pipeline developed by Singhal (2013; https://github.com/MVZSEQ) to remove duplicates, merge overlapping reads, and trim poor

quality bases and adaptor sequences. We then assembled the clean reads for each sample by identifying, for each exon, sequencing reads that are homologous to *Anolis* based on blastx (Altschul et al. 1990). This was followed by a de novo assembly of the identified reads using Velvet v.3 (Zerbino & Birney 2008) and cap3 (Huang & Madan 1999), and a subsequent trimming of the contigs to the exonic regions using Exonerate v.2.2.0 (Slater et al. 2005).

We mapped, with bowtie v.2.2.2 (Langmead & Salzberg 2012), the clean reads back to the assembled target sequences for each sample as its own reference and the resulting SAM files were processed using SAMTOOLS v.0.1.19 (Li et al. 2009). To identify heterozygous sites, we masked sites with a low-quality genotype score (GQ < 20) and phased overlapping individual sequencing reads within loci ("read-backed phasing") using GATK v.3.3 (McKenna et al. 2010). The resulting output were two haplotype sequences for each locus in each sample (labelled as h0 and h1), where for the unresolved heterozygous sites the reference allele was assigned to the h0 haplotype and the alternate allele to h1. The scripts used for processing sequencing reads and processed data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.34274. To identify the samples of best quality and that were most informative, we assessed sequencing success and assembly quality by calculating coverage and the number of exons recovered per sample, and ultimately selected the samples with more than 1900 loci.

We used EAPhy v. 1.1 (Blom 2015) to align the exons with MUSCLE v.3.8.31 (Edgar 2004) and review the quality of these alignments based on several filter criteria; tasks performed included removing missing data from the ends of sequences, checking for translation of amino acids, and exporting both single-locus and concatenated alignments in different formats and with different proportions of missing data. For subsequent sequence-based analyses, we extracted loci from a general alignment with the 228 samples plus the outgroup. For the SNP-based analyses the exons were realigned, based on subsampling for downstream analyses, to recover the largest possible number of SNPs, and we randomly extracted one biallelic SNP per locus from sites that were not missing data for most of the samples. For further information about missing data for each analysis see Appendix S5.2, Supplementary material).

5.2.2 Recent admixture inference

To visualize potential recent hybridization among all six species, we performed an exploratory clustering analysis including all individuals using a principal coordinate analysis (PCoA) based on Euclidean genetic distances, as implemented in the package ADE4 (Dray & Dufour 2007) run in R v.3.3.2 (R Core Team 2016). We also tested for the presence of admixture between any pair of species using STRUCTURE v.2.3.4 (Pritchard et al. 2000). We conducted a total of 15 pairwise comparisons to estimate admixture between each pair of species using K = 2 (see Appendix S5.2, Supplementary material for detailed methods). These analyses were performed using the admixture model with independent allele frequen-

cies, and 20 simulation replicates, each with a burn-in of 50,000 iterations followed by sampling for 500,000 generations. After verifying MCMC convergence between independent runs, we ran CLUMPP v.1.1.2 (Jakobsson & Rosenberg 2007) to align them. To assess the effect of species' genetic structure on the membership coefficients, we compared these results with additional pairwise comparisons using one lineage per species for both the Kimberley and the Top End.

5.2.3 Exploring conflicting signals between exons

We initially explored patterns of reticulation among all species with the Neighbor-Net algorithm in SplitsTree4 v.4.14.6 (Huson & Bryant 2006). This is an unrooted distance-based method that can reveal conflicts due to incompatible phylogenetic signals in a concatenated dataset. For this analysis we used a concatenated alignment of 1805 loci with one haplotype from each individual.

To observe species tree support for the six species topology accounting for the presence of ILS, we ran ASTRAL-II v.4.11.1 (Mirarab & Warnow 2015), a summary-coalescent method that finds the species tree that shares the maximum number of induced quartet trees with gene trees. An advantage of this method is that it can compute support for species tree branches based on both gene tree bootstraps and local posterior probabilities derived from gene tree quartet frequencies (Sayyari & Mirarab 2016). The local posterior probabilities are calculated by assuming that for each internal branch of a given species tree three topologies are possible around that branch, and a quartet support is computed for each of those three alternatives (Sayyari & Mirarab 2016). For this analysis, we extracted a total of 1592 individual exons from the main alignment using one haplotype from each of the 18 lineages and the *C. rubrigularis* outgroup. For each exon, we ran JModeltest v.2.1.10 (Darriba et al. 2012) to select the best-fit substitution model, and used RAxML v.8.2.9 (Stamatakis 2014) to estimate the best tree out of ten search replicates, with branch support based on 100 bootstrap replicates.

5.2.4 Introgression between species and lineages

We searched for signatures of introgression between each pair of non-sister species by considering all combinations of possible asymmetric trees and using Patterson's D statistic (Green et al. 2010; Durand et al. 2011), also known as the ABBA-BABA test. This method considers an asymmetric four-taxon tree (((P1, P2), P3), O), where O is the outgroup, P1 and P2 are a sister group, and P3 is sister to (P1, P2) that we want to test for introgression with either P1 or P2. This test uses only biallelic SNPs, for which the outgroup allele is expressed as "A" (i.e. ancestral) and the alternative allele as "B". The statistic tests if the numbers of derived alleles shared between P1 and P3 and between P2 and P3 are similar, a pattern expected from ILS (see Appendix S5.3, Supplementary material).

We performed this approach at both the species and lineage levels, the latter to address the goal of comparing the two regions that experienced different climatic fluctuations

in the past, the Kimberley and Top End. To do this, we started by subsampling our dataset to two lineages per species, one of the lineages mainly distributed in the Kimberley and the other mostly in the western Top End. We also subsampled the number of individuals to a maximum of five samples per lineage, but evenly covering the geographic distribution of the lineages and including locations where more than one species was collected.

We repeated the alignments with the selected subset of samples and extracted all segregating sites using EAPhy. To estimate Patterson's D statistic between species for as many combinations of asymmetric four-taxon trees (20 analyses; Fig. 5.3), we calculated ABBA and BABA patterns from estimating the allele frequencies in each taxon for each biallelic SNP across all loci, so the statistic could be estimated similarly to the equation (2) of Durand et al. (2011). We used this output to calculate D with the dstat mode implemented in dfoil (Pease & Hahn 2015) and a χ^2 goodness-of-fit test to evaluate significance ($\alpha = 0.01$).

To test for introgression at the lineage-level and detect if the patterns would be different between the Kimberley and the Top End lineages, we used the five-taxon ABBA-BABA D_{FOIL} statistic developed by Pease & Hahn (2015). This approach is an extension of the four-taxon ABBA-BABA test and is also based on biallelic sites but uses a five-taxon symmetric phylogeny with two pairs of ingroup taxa and an outgroup taxon, (((P1,P2),(P3,P4,)),O), where P3 and P4 diverged no later than P1 and P2 did. The main advantage of this method is that it allows inferences about the presence and direction of introgression by using a system of four D statistic-like components (referred as D_{FO}, D_{IL}, D_{FI} and D_{OL}). We performed the five-taxon ABBA-BABA test in dfoil based on allele frequencies, including the two lineages per species, and in a pairwise fashion, for a total of 15 comparisons; each component significance was tested by a χ^2 goodness-of-fit test with a cut-off of $P < 0.01$ (Table 5.1; Appendix S5.3, Supplementary material).

For all analyses, we resampled exons with replacement and generated 1000 bootstrap pseudoreplicates SNP datasets to obtain 95% confidence intervals. Analyses were conducted and plotted using customized R scripts with the seqinr package (Charif & Lobry 2007) and packages within tidyverse (Wickham 2017).

5.2.5 Reticulation from phylogenetic networks

Using the sequences from the same dataset as in the ABBA-BABA tests, we applied another statistical approach based on gene tree discordances, to infer phylogenetic networks while accounting for ILS. SNaQ - Species Networks applying Quartets - (Solís-Lemus & Ané 2016) uses pseudolikelihood approximations to the full model likelihood under a coalescent-based model and is implemented in the package PhyloNetworks v.0.3.0 (Solís-Lemus et al. 2017) in Julia (Bezanson et al. 2017). This allows us to include both lineages from each of the species studied in a single analysis. Compared with related MSNC methods (such as the ones implemented in PhyloNet, Than, Ruths & Nakhleh, 2008), the algorithm in PhyloNetworks is fast (Hejase & Liu 2016), but estimating phylogenetic networks from multilocus data is computationally intensive with increasing number of tips and we therefore ran

five replicate analyses, each using a different single haplotype per lineage (Appendix S5.2, Supplementary material). PhyloNetworks does not estimate gene trees directly, but instead requires a table of quartet concordance factors (CFs) observed from multi-locus data as input. Therefore, for each replicate analysis, we ran MrBayes v. 3.2.3 (Ronquist et al. 2012) for each locus separately and the respective posterior distribution of trees was used as input for BUCKy v.1.4.4 (Ané et al. 2007; Larget et al. 2010) to estimate the quartet CFs. This is the preferred strategy for estimating CFs to be used in PhyloNetworks, because it estimates true gene tree conflict, discards conflict due to gene tree error and provides 95% credibility intervals, with CFs estimates least influenced by genes with large tree uncertainty (Solís-Lemus & Ané 2016). This strategy was implemented in the pipeline developed by Stenz et al. (2015), available at https://github.com/nstenz/TICR. In PhyloNetworks, we estimated for each replicate a network from one to four reticulations with the best pseudolikelihood after 20 runs using the SNaQ function. Subsequently, we performed 100 bootstrap replicates on the best network, where for each replicate with 10 runs, 30% started from the estimated network and 70% from the species tree.

Additionally, using the sites with synonymous substitutions of the same data set used in the ABBA-BABA tests, we estimated nucleotide diversity (pi), average sequence divergence per site (Dxy; Nei, 1987) and net average sequence divergence per site (Da; Nei, 1987) with the R package PopGenome (Pfeifer et al. 2014) to characterize genetic diversity and divergence within and among lineages and species.

5.3 Results

Across all samples sequenced in the current and past studies, we used the 229 individuals with approximately >2000 loci from *C. gracilis*, *C. munda*, *C. rufilatus*, *C. amax*, *C. triacantha (s. l.)*, *C. johnstonei (s. l.)*, and *C. rubrigularis* (outgroup). On average, the dataset had 2811 loci, 128x coverage depth, and a proportion of heterozygous sites per individual of $3.47x10^{-3}$ (Appendix S5.2, Supplementary material). It is worth noting that, based on previous studies, all individuals except one had concordance for nDNA and mtDNA species identity (data not shown). For silent sites, estimated net (Da) divergences among species (estimated at lineage level to avoid confounding effects of deep phylogeographic structure) ranged from ~ 0.4% to ~1%; the range of corresponding average absolute divergences (Dxy) was similar - ~0.6% to 1.1% (Appendix S5.4, Supplementary material).

5.3.1 Recent admixture

For exploratory analyses, we examined PCoA results including all six species (Appendix S5.5, Supplementary material). Besides *C. johnstonei (s. l.)* and *C. triacantha (s. l.)* which overlap at this scale, all but a few individuals cluster tightly within nominal species; these generally tight genotypic clusters imply a low level of recent admixture overall.

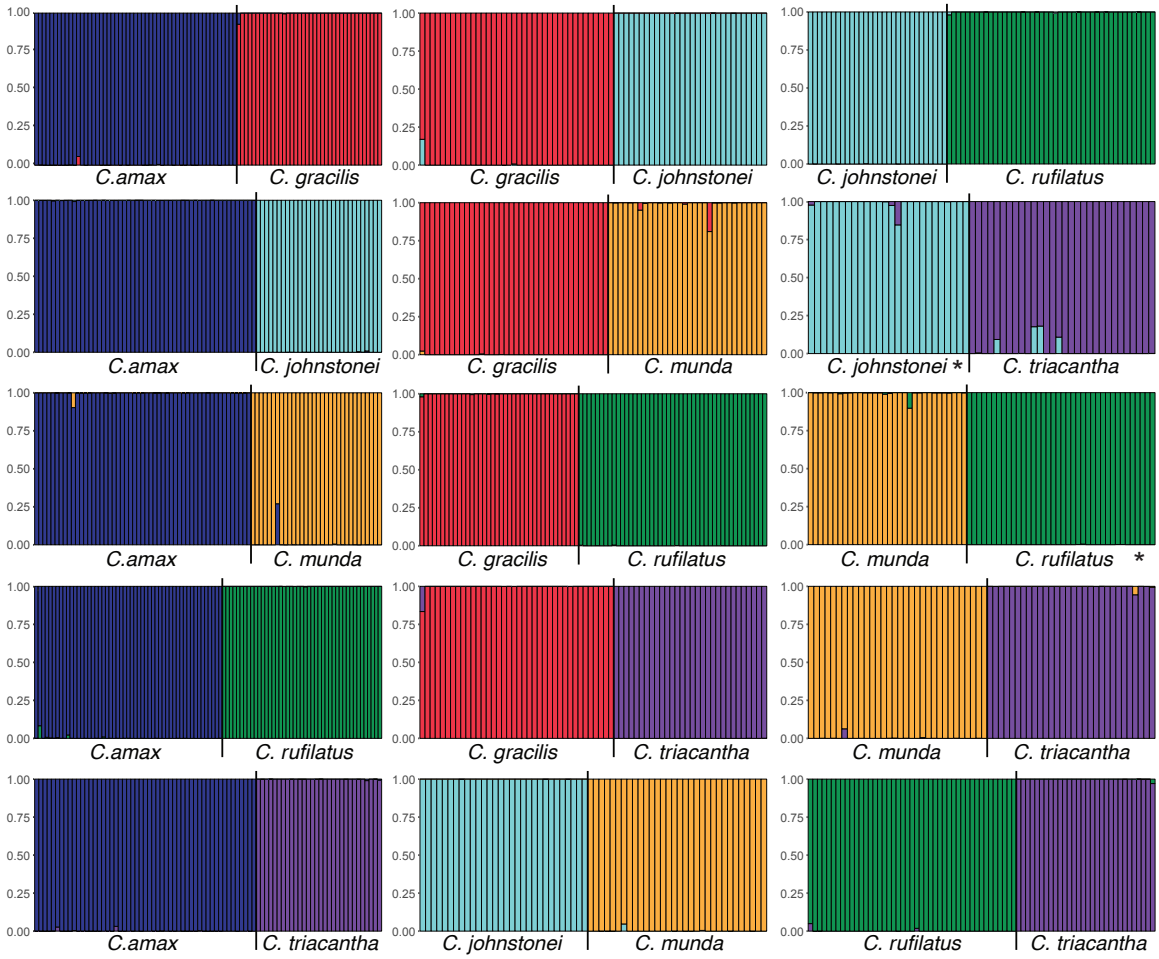


Figure 5.2 Bar plots of STRUCTURE results for pairwise comparisons between the six studied species. Both *C. johnstonei* and *C. triacantha* are used here in *sensu lato*.
* The same individuals were used across all comparisons, except in those between sister species (*C. johnstonei* vs. *C. triacantha* and *C. rufilatus* vs. *C. munda*) where individuals representing highly divergent lineages (*C. insularis* within *C. johnstonei* and ECI of *C. rufilatus*) were removed because this genetic structure affected the inference of admixture.

More quantitatively, we then investigated recent admixture using pairwise clustering across all species with the model-based admixture analysis (K = 2; STRUCTURE). Overall, it also suggested little interspecific admixture (membership below 90% consistent across replicate runs) (Fig. 5.2). The few exceptions were mostly individuals evidently admixed between the closely related species *C. triacantha (s. l.)* and *C. johnstonei*; other admixed individuals included a *C. munda* individual admixed with *C. gracilis* and *C. rufilatus*, a *C. gracilis* individual admixed with *C. triacantha* and *C. johnstonei*, and a *C. munda* individual highly admixed with *C. amax*. When considering samples from the Kimberley and Top End separately, there were indications of slight admixture in some samples from the former region, but none in the latter region (Appendix S5.6 and S5.7, Supplementary material).

5.3.2 Exploring gene tree conflict

To explore evidence of reticulation between the six species and of conflicting phylogenetic signals among exons, we used both the Neighbor-Net algorithm, which ignores coalescent processes in the species tree, and the coalescent-based species tree estimation method ASTRAL-II that takes into account the presence of ILS when estimating species relationships. Although sensitive to introgression (Solís-Lemus et al. 2016), ASTRAL is an effective way of evaluating gene tree discordance at nodes in a tree. The distance method results suggested potential reticulation between *C. gracilis* and the *C. munda* + *C. rufilatus* clade, and also within the *C. johnstonei (s. l.)* and *C. triacantha (s. l.)* clade (Fig. 5.1b). While the species tree (Fig. 5.1c) identifies a species tree topology congruent with previous studies (Bragg et al. 2018), most nodes have substantial gene tree discordance, with similar frequencies of the two minor topologies as expected for ILS alone.

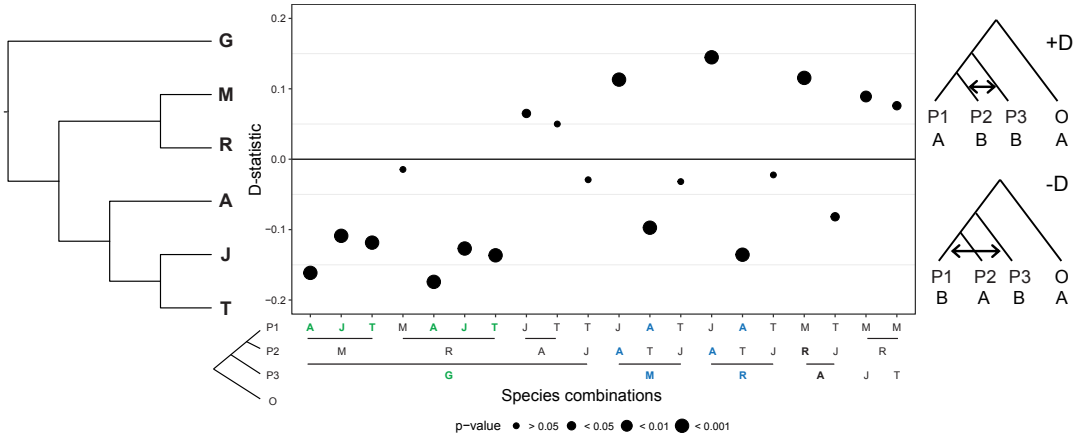


Figure 5.3 Patterson's D-statistic estimates from four-taxon ABBA/BABA tests for the different possible species combinations and species tree configurations. The size of the black circles in the central plot reflect the observed p-value according to four arbitrarily defined ranges of p-values; we used a significance level of 0.01. A positive D statistic indicates potential introgression between species P2 and P3, while a negative D suggests introgression between P1 and P3. Green correspond to tests supporting evidence for G-AJT introgression and blue for A-MR.

5.3.3 Ancient introgression

We wanted to investigate introgression at two levels: between species before intraspecific lineage divergence; and between broadly sympatric lineages from different species, focusing on one lineage occurring mostly in the Kimberley (k) vs a lineage occurring mostly in the Top End (t) region. For this, we identified 20 possible combinations of sets of three species to conduct ABBA-BABA tests (Fig. 5.3, Appendix S5.3 and S5.8, Supplementary material), and then classified the species with evidence of introgression: a significant positive D suggested introgression between species P2 and P3, and a significant negative D introgression between P1 and P3. From the species-level tests, we found strong evidence of introgression between *C. gracilis* and each of *C. amax*, *C. johnstonei* and *C. triacantha*, which could mean introgression involving the common ancestor of these species (i.e. G-AJT). This

was confirmed by using either *C. rufilatus* or *C. munda* as the P2 species. Another strongly supported result was between *C. amax* with both *C. rufilatus* and *C. munda* (A-RM, and using either *C. johnstonei (s. l.)* or *C. triacantha (s. l.)* as the P2 species), with the D values being higher in the tests for the first comparison. Accordingly, we found evidence for introgression between *C. amax* and *C. rufilatus* in the MRA test (P1P2P3 as in Fig. 5.3). A lower magnitude D was also detected between *C. rufilatus* and *C. johnstonei*. The exons bootstrap analyses showed relatively wide 95% confidence intervals with overall similar patterns, except for the TJM combination (Appendix S5.9, Supplementary material).

Table 5.1 D_{FOIL} results for species comparisons at the lineage-level.

P1	P2	P3	P4	D_{FO}		D_{IL}		D_{FI}		D_{OL}		Introgression
				D' stat	sign	D' stat	sign	D' stat	sign	D' stat	sign	
Gk	Gt	Mk	Mt	0.085	+	0.085	+	-0.019	0	-0.018	0	G - Mk
Gk	Gt	Rk	Rt	0.002	0	0.003	0	-0.016	0	-0.016	0	-
Gk	Gt	Ak	At	-0.001	0	-0.002	0	-0.019	0	-0.021	0	-
Gk	Gt	Jk	Ji	-0.080	-	-0.081	-	-0.010	0	-0.013	0	G - Ji
Gk	Gt	Tk	Tt	-0.053	-	-0.052	-	-0.021	0	-0.020	0	G - Tt
Mk	Mt	Ak	At	0.005	0	0.005	0	0.038	+	0.039	+	-
Mk	Mt	Rk	Rt	0.004	0	0.003	0	0.021	0	0.020	0	-
Mk	Mt	Jk	Ji	-0.077	-	-0.078	-	0.036	+	0.034	+	-
Mk	Mt	Tk	Tt	-0.052	-	-0.051	-	0.038	+	0.039	+	-
Rk	Rt	Ak	At	0.000	0	0.002	0	-0.004	0	-0.002	0	-
Rk	Rt	Jk	Ji	-0.082	-	-0.075	-	0.005	0	0.014	0	R - Ji
Tk	Tt	Rk	Rt	0.006	0	0.005	0	-0.051	-	-0.052	-	-
Jk	Ji	Ak	At	-0.007	0	-0.002	0	-0.081	-	-0.075	-	-
Tk	Tt	Ak	At	0.001	0	0.001	0	-0.046	-	-0.047	-	-
Tk	Tt	Jk	Ji	-0.091	-	-0.078	-	-0.066	-	-0.048	-	-

We further used the five-taxon ABBA-BABA approach to test simultaneously for the presence and direction of introgression between lineages from different species in 15 pairwise comparisons that confirmed to the required balanced topology (Appendix S5.10, Supplementary material). All significant results suggest evidence of ancestral introgression (Table 5.1; Appendix S5.10, Supplementary material), but we are unable to determine the directionality in this situation (Appendix S5.3, Supplementary material). The D_{FOIL} approach indicated introgression between *C. gracilis* (Gk+Gt) and lineages from *C. munda* (Mk), *C. triacantha* (Tt) and *C. insularis* (Ji). We used Ji samples to include two lineages from *C. johnstonei (s. l.)*, thus allowing to check for introgression in *C. johnstonei (s. s.)*, and we found support for introgression between Ji and *C. munda* (Mk+Mt) and also with *C. rufilatus* (Rk+Rt). The bootstrap results were similar to the original analysis and the confidence intervals were narrower compared to the four-taxon ABBA-BABA bootstrap analysis (Supplemen-

tary material). Some results might not be conclusive (absence of introgression between lineages within M-T and R-T), considering the short distance between P1/P2 to P3/P4 taxa (Appendix S5.10, Supplementary material).

We also employed an alternative approach to infer reticulation events using full sequence data and an explicit phylogenetic network method that accounts for ILS using the MSNC model. Because it is computationally intensive, we used one individual per lineage but ran five different replicates of samples to use a dataset more similar to that used in the ABBA-BABA analyses. The most consistent result across all five replicates supports one introgression event between *C. amax* and the *C. munda* + *C. rufilatus* clade (A-MR) (estimated inheritance probability of 0.35), and with a topology placing *C. gracilis*, rather than C.munda - *C. rufilatus*, as most closely related to the *C. amax* + *C. johnstonei* + *C. triacantha* clade (G-AJT) (Fig. 5.4A,B). This result shown is for the best model based on four of the five replicates of samples (excluding models with a reticulation involving the outgroup). For the fifth replicate, the best model had the same topology as the species tree with a reticulation event between *C. gracilis* and the *C. johnstonei (s. l.)* + *C. triacantha (s. s.)* clade (G-TJ), but in this network *C. insularis* (Ji) was not sister species to *C. johnstonei (s. s.)* (Fig. 5.4C,D; Appendix S5.12, Supplementary material). This network was also one of the best models across the other four replicates, albeit with a lower pseudolikelihood.

The bootstrap networks can provide information on the support for branches, including hybrid edges and direction of reticulation events. For the most consistent network, the one with A-MR reticulation, there was strong support for all branches, being *C. gracilis* the least well supported (Fig. 5.4A). The hybrid edge was recovered in 87% of the replicates in which the direction of introgression was from *C. amax* to the MR clade (Fig. 5.4B, see also Appendix S5.13, Supplementary material). By contrast, in the second network (Fig. 5.4C,D), despite having the expected topology, the support for the node connecting AJT and MR was very low, and the bootstrap support for the G-JT reticulation was also very low. The Figure 5.4D that represents the origin and target of gene flow indicates low support across different alternatives, with the originally estimated reticulation showing 33% of support for *C. gracilis* as the target (41% as the origin) and 18% for JT as the origin of gene flow.

Considering the PhyloNetwork suggestion that *C. gracilis* is not the more divergent taxon, the order of species in the ABBA-BABA results could be incorrect, but additional tests showed that the expected species tree topology is more appropriate to perform the ABBA-BABA tests (Appendix S5.14, Supplementary material). This is further supported by the larger Dxy and Da values between *C. gracilis* and the other species (Appendix S5.4, Supplementary material).

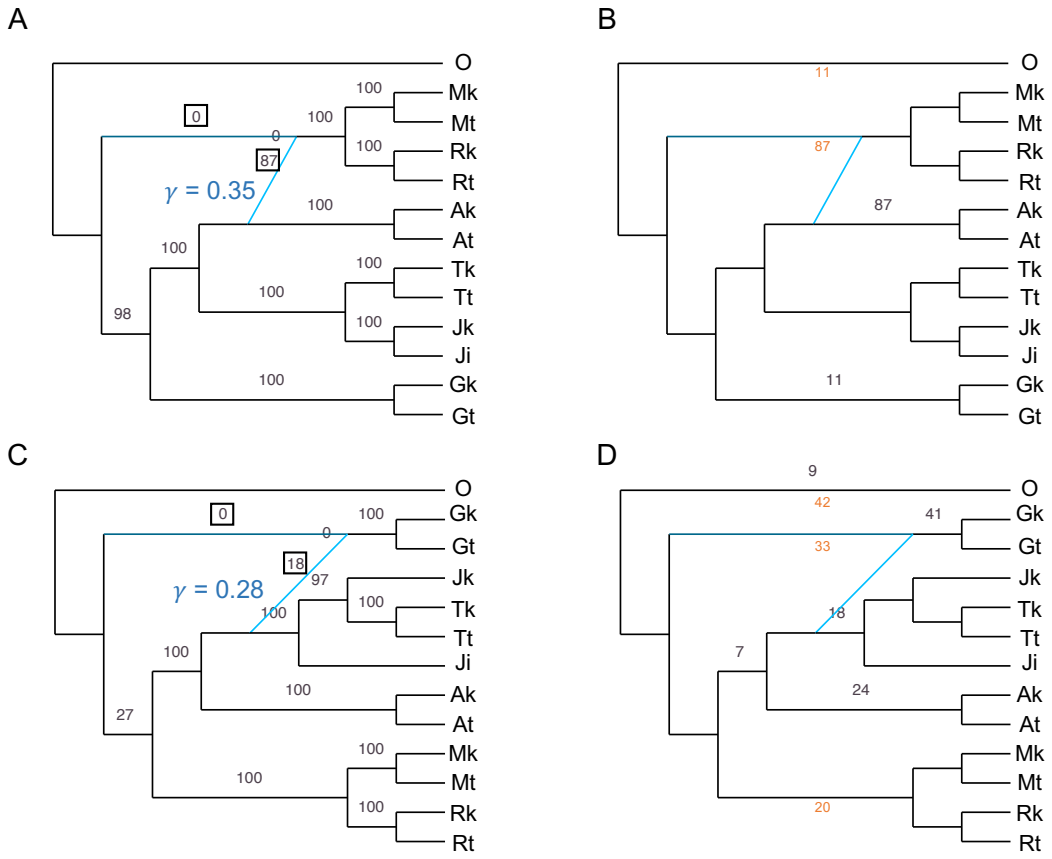


Figure 5.4 PhyloNetworks bootstrap networks for replicates 1 (A and B) and 5 (C and D). The left panels show bootstrap values for branches; inside the squares are the bootstrap support for the hybrid edges (i.e. percentage of bootstrap trees in which the same sister clade is grouped with the hybrid clade) and the hybrid node. The number in blue is the inheritance probability, γ, which represents the proportion of genes contributed by each parental population to a given hybrid node. The right panels give the bootstrap values depending on whether a given lineage is the origin (above the edge) or target of gene flow (below the edge and in orange).

5.4 Discussion

This study uses new and published exon capture datasets to explore the potential role of introgression in the evolution of sympatric species from the Australian Monsoonal Tropics (AMT). Despite broad overlap in geographic distributions between the AMT *Carlia* species, signatures of recent admixture between different species seem to be negligible. Nonetheless, we uncovered several ancestral introgression events (summary in Fig. 5.5).

5.4.1 Near absence of recent introgression among species

Given that these species are closely related and often sympatric, it is reasonable to expect some evidence of recent hybridization, perhaps declining with the magnitude of genetic divergence. This expectation also follows from the observation that net divergences among the AMT *Carlia* are in the range of transition between low and high probabilities of gene flow in comparative studies (Roux et al. 2016) and where, in related species, some

hybridization (but not introgression) is observed among morphologically cryptic lineages (Singhal & Moritz 2013; Singhal & Bi 2017). Yet, for the AMT species examined here, there is very little evidence of recent admixture from clustering (STRUCTURE) analyses. This implies stronger reproductive isolation among these species than expected based on genomic divergence alone. Given the marked differences in male breeding colours of the "rainbow skinks", we suggest that the evidently strong reproductive isolation among these taxa could reflect colour-based assortative mating, as observed from experimental studies in red- vs blue-throated *Carlia* from the rainforests of the east coast (Dolman 2008). Although colour variations within the AMT *Carlia* skinks are subtler (Dolman & Stuart-Fox 2010), this hypothesis is amenable to experimental test, contrasting female response to males of the same or different species and across a range of divergence in colours differing in the extend of colour.

5.4.2 Ancestral introgression between species

Support of ancestral introgression between species comes from both the ABBA-BABA tests and the phylogenetic networks results that consistently suggest introgression between *C. amax* and the *C. munda* + *C. rufilatus* clade (A-MR). The PhyloNetwork analysis also suggests that the direction of gene flow is from *C. amax* to the MR clade, with around a 0.35 proportion of genes contributed by each parental population to the MR hybrid node. The ABBA-BABA tests also infer a reticulation event between *C. gracilis* and likely the ancestor of *C. amax*, *C. johnstonei* and *C. triacantha* (G-AJT). These two events are further supported with additional ABBA-BABA analyses (Appendix S5.8, Supplementary material) in which sister taxa were lumped together (i.e., M+R, A+J+T and J+T). An exception was for the A-(M+R)-G test that unexpectedly supported introgression between *C. amax* and *C. gracilis*, rather than A-MR as inferred from phylogenetic networks.

The phylogenetic network analyses also yielded a topology that differs from the concatenated and ASTRAL species trees, with the munda-rufilatus pair being the most divergent group rather than *C. gracilis*. We also note that the node uniting MR with (AJT) in the ASTRAL tree has extensive gene tree discordance. Given the known sensitivity of concatenation and ASTRAL methods to introgression (Solís-Lemus et al. 2016), and PhyloNetworks advantage to infer an appropriate topology while estimating introgression (Solís-Lemus & Ané 2016), it could be that the dominant topology from the network analyses is correct. This issue could be further explored using full Bayesian network methods that directly incorporate ILS and introgression while simultaneously estimating the gene trees (Wen et al. 2016; Zhang et al. 2017). However, at present these are computationally demanding for data sets of this size in number of loci and taxa. Irrespective of the topology, one interesting observation can be made – whereas historical introgression is supported among distantly related taxa or clades, there is no such evidence between *C. amax* and the more closely related triacantha-johnstonei clade. While it is tempting to speculate about whether morphologically or ecological similarities are related to past introgression events, the marked differences in traits between,

for example, sister taxa (i.e. M vs R, and T vs J; see Potter et al. 2017; Afonso Silva et al. 2017a) and their relatively old divergence, would render estimation of ancestral states uncertain.

Two possible scenarios are likely considering the overall results (Fig. 5.5): a first scenario with two reticulation events – G-AJT and A-MR, which are supported by much of the ABBA-BABA tests. In this scenario, *C. gracilis* is the more divergent species within this clade, and *C. amax* diverged from JT and before the MR split. The second scenario is more consistent with the PhyloNetwork analysis, allowing for an introgression event (A-MR) and identifying the MR clade as more ancestral while *C. gracilis* is more closely related to the AJT clade. Considering that the species tree in Bragg et al. (2018) also supported the position of *C. gracilis*, it is possible that these relatively unclear topology results might reflect divergence with gene flow between *C. gracilis* and the MR ancestor, with recurrent events of admixture between species.

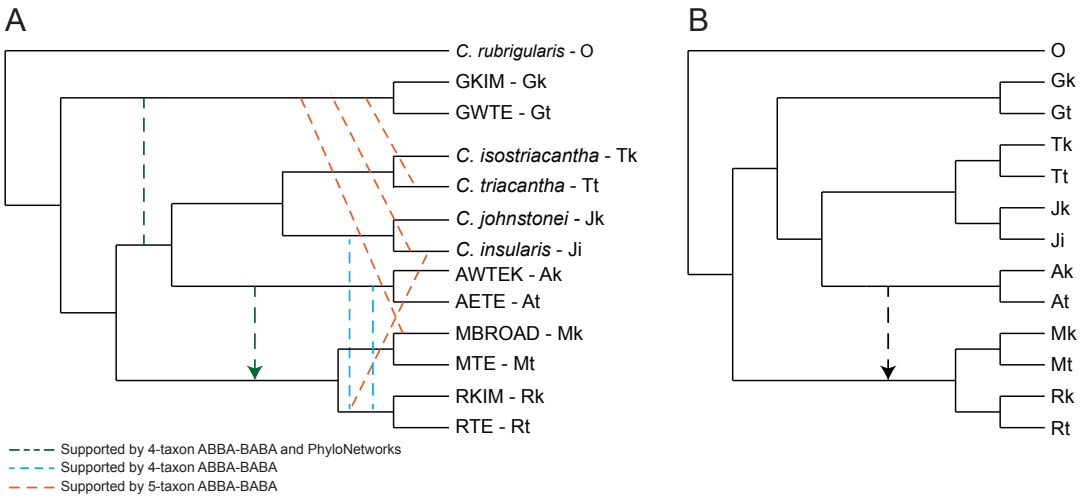


Figure 5.5 Summary of the main results from the ABBA-BABA and PhyloNetworks analyses. Panels A and B show the trees with either *C. gracilis* or MR as the most basal ingroup clade, respectively. The dashed lines represent potential introgression events inferred by the different methods and are color-coded as per the legend in the figure. The dashed line with arrow highlights the well-supported direction of introgression between *C. amax* and the MR clade in PhyloNetworks analyses based on either of the two topologies.

5.4.3 Climate dynamics, range shifting and opportunity for hybridization

Paleoclimate models and analyses of historical demography point to a more dynamic and disparate history of population fluctuations across the Kimberley than the Top End (Potter et al. 2016, 2017; Afonso Silva et al. 2017a). As habitat instability, range fluctuations and range invasions can promote hybridisation and introgression (Anderson 1948; Excoffier et al. 2009; Canestrelli et al. 2016), we expected more signals of current and historical introgression in the Kimberley than the Top End. While signals of recent introgression are sparse, we find more evidence of ancestral introgression between taxa from the Kimberley, with the

two instances involving the species *C. gracilis*, with *C. munda* BROAD lineage and *C. insularis*, and a third instance between the latter species and *C. rufilatus*. This compared with a single introgression event in the Top End, between *C. gracilis* and *C. triacantha (s. s.)* (which extends into the central deserts).

Although introgression between lineages that only overlap in either region was not detected, the above findings could suggest some role of range expansion in hybridization, considering that all three cases involve taxa with different distributions when compared to the other tested lineages: either very restricted (*C. insularis*, endemic to small land-bridge Kimberley islands) or very widespread (*C. triacantha* and *C. munda* BROAD, each respectively occupying the Top End and Kimberley, and both also more broadly distributed into the arid regions of Central Australia).

5.4.4 Methodological limitations

Our analyses sought to investigate both recent and historical introgression using multiple approaches for each of the two levels. The recent proliferation of methods for detecting ancestral introgression is a very welcome addition to our inference tools, especially given increasing evidence for the prevalence and evolutionary significance of gene flow among animal species (Mallet 2005; Abbott et al. 2013; Mallet et al. 2016). While the methods employed here – ABBA-BABA (four and five-taxon) and MSNC-based (PhyloNetworks) methods – did reveal congruent signals of introgression, there were also some idiosyncratic results. The reasons for this are not clear; the methods differ in the input data (SNPs vs sequences), topologies used (ABBA-BABA vs D_{FOIL}) and in the underlying assumptions. We also recognize the potential for false signals of introgression arising from structured ancestral populations (Durand et al. 2011; Eriksson & Manica 2012) and so have focused our discussion on results supported by multiple methods.

More generally, there is a need for more simulation-based analyses of power and sensitivity to model violations, such as genetic structure, differences in substitution rate and divergence levels between analysed species, in the application and comparison of ABBA-BABA and MSNC methods to different scenarios of recent and ancient admixture. Additionally, we expect that further methodological developments and richer datasets, especially whole genome sequencing, will provide better resolution to explore introgression across the genome considering the ancient nature of many of these patterns.

5.4.5 Introgression role in species evolution

A key result from the above analyses is that the simple prediction that there would be more evidence of recent hybridization, and of recent or ancestral introgression, among closely related than distantly related species is not supported. This contrasts with comparative studies of related, and similarly divergent, but morphologically cryptic taxa from the rainforests of eastern Australia (Singhal & Moritz 2013; Singhal & Bi 2017). Rather, for the AMT skinks, we found evidence for older introgression events. This contrast is nevertheless

perhaps not surprising and serves to remind us that reproductive isolation accrues at different time scales (Gavrilets 2003; Coyne & Orr 2004; Sobel et al. 2010), with different rates and outcomes for morphology and genomic divergence, and as a result of the many speciation processes that are in play.

In fact, the role of gene flow across different scales within the *Carlia* species evolutionary histories suggests some parallels with examples from the well-studied African Lakes cichlids. In this famous example of adaptive radiation, historical introgression has been suggested to fuel rapid speciation across multiple groups (Salzburger et al. 2002; Genner & Turner 2012; Meier et al. 2017) and to be an important force in rapid adaptation to changing environments (Seehausen 2004). Similarly, it has been shown that introgression occurred throughout the evolutionary history of lamprologines cichlids (Gante et al. 2016). Despite much more conservative morphological variation across *Carlia*, the fact that introgression patterns are observed broadly across other groups with sympatric species like the cichlids, suggests a need to explore introgression more broadly across the rapid radiation of the rainbow skinks, and to what extent this might have coincided with periods of environmental instability.

Studying how introgression has occurred in the past across different levels of divergence is currently more of a possibility than ever before and has great implications not only for the understanding of the speciation process, but also the impact of hybridization as species respond to current climatic change.

5.5 References

Abbott R, Albach D, Ansell S et al. (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Afonso Silva AC, Bragg JG, Potter S et al. (2017a) Tropical specialist vs. climate generalist: diversification and demographic history of sister species of Carlia skinks from northwestern Australia. *Molecular Ecology*, **26**, 4045–4058.

Afonso Silva AC, Santos N, Ogilvie HA, Moritz C (2017b) Validation and description of two new northwestern Australian Rainbow skinks with multispecies coalescent methods and morphology. *PeerJ*, **5**, e3724.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Anderson E (1948) Hybridization of the Habitat. *Evolution*, **2**, 1–9.

Anderson E, Stebbins Jr. GL (1954) Hybridization as an evolutionary stimulus. *Evolution*, **8**, 378–388.

Ané C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution*, **24**, 412–426.

Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. *Trends in Ecology & Evolution*, **25**, 530–536.

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A fresh approach to numerical computing. *SIAM Review*, **59**, 65–98.

Bi K, Vanderpool D, Singhal S et al. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, **13**, 403.

Bragg JG, Potter S, Bi K, Moritz C (2016) Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources*, **16**, 1059–1068.

Brawand D, Wagner CE, Li YI et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375.

Canestrelli D, Porretta D, Lowe WH et al. (2016) The tangled evolutionary legacies of range expansion and hybridization. *Trends in Ecology & Evolution*, **31**, 677–688.

Charif, D. & Lobry, J.R. (2007) SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *Structural approaches to sequence evolution: molecules, networks, populations* (ed. by U. Bastolla, M. Porto, H.E. Roman and M. Vendruscolo), Vendruscolo), pp. 207–232. Springer, New York.

Cogger H (2014) *Reptiles and amphibians of Australia*. New Holland, Sydney.

Coyne JA, Orr HA (2004) *Speciation.* Sinauer & Associates, Sunderland, Massachusetts.

Cui R, Schumer M, Kruesi K et al. (2013) Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, **67**, 2166–2179.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–340.

Deutsch CA, Tewksbury JJ, Huey RB et al. (2008) Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences*, **105**, 6668–72.

Dolman G (2008) Evidence for differential assortative female preference in association with refugial isolation of rainbow skinks in Australia's tropical rainforests. *PloS one*, **3**, e3499.

Dolman G, Hugall AF (2008) Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Molecular Phylogenetics and Evolution*, **49**, 782–94.

Dolman G, Stuart-Fox D (2010) Processes driving male breeding colour and ecomorphological diversification in rainbow skinks: a phylogenetic comparative test. *Evolutionary Ecology*, **24**, 97–113.

Dray S, Dufour AB (2007) The ade4 Package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using radseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–7.

Edwards S V., Potter S, Schmitt CJ, Bragg JG, Moritz C (2016) Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences*, **113**, 8025–8032.

Eriksson A, Manica A (2012) Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proceedings of the National Academy of Sciences*, **109**, 13956–13960.

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Fontaine MC, Pease JB, Steele A et al. (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, **347**, 1258524.

Gante HF, Matschiner M, Malmstrøm M et al. (2016) Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Molecular Ecology*, **25**, 6143–6161.

Gavrilets S (2003) Models of speciation: what have we learned in 30 years? *Evolution*, **57**, 2197–2215.

Genner MJ, Turner GF (2012) Ancient hybridization and phenotypic novelty within lake malawi's cichlid fish radiation. *Molecular Biology and Evolution*, **29**, 195–206.

Gompert Z, Mandeville EG, Buerkle CA (2017) Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, **48**, 207–229.

Green RE, Krause J, Briggs AW et al. (2010) A draft sequence of the Neanderthal genome. *Science*, **328**, 710–722.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795–809.

Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.

Hejase HA, Liu KJ (2016) A scalability study of phylogenetic network inference methods using empirical datasets and simulations involving a single reticulation. *BMC Bioinformatics*, **17**, 422.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome research*, **9**, 868–77.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–67.

Ivory SJ, Blome MW, King JW et al. (2016) Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *Proceedings of the National Academy of Sciences*, **113**, 11895–11900.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

James C, Shine R (1988) Life-history strategies of Australian lizards: a comparison between the tropics and the temperate zone. *Oecologia*, **75**, 307–316.

Kumar V, Lammers F, Bidon T et al. (2017) The evolutionary history of bears is characterized by gene flow across species. *Scientific Reports*, **7**, 1–10.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Larget BR, Kotha SK, Dewey CN, Ané C (2010) BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, **26**, 2910–2911.

Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–9.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–37.

Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays*, **38**, 140–149.

Marques JP, Farelo L, Vilela J et al. (2017) Range expansion underlies historical introgressive hybridization in the Iberian hare. *Scientific Reports*, **7**, 40788.

McKenna A, Hanna M, Banks E et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**, 1297–303.

Meier JI, Marques DA, Mwaiko S et al. (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, **8**, 14363.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **5**, 1-10.

Mirarab S, Warnow T (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, **31**, i44–i52.

Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Nosil P (2012) *Ecological speciation*. Oxford University Press.

Ottenburghs J, Megens HJ, Kraus RHS et al. (2017) A history of hybrids? Genomic patterns of introgression in the True Geese. *BMC Evolutionary Biology*, **17**, 1–14.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337–2360.

Pease JB, Hahn MW (2015) detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, **64**, 651–662.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ (2014) PopGenome: an efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31, 1929–1936.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, **25**, 1367–1380.

Potter S, Xue AT, Bragg JG et al. (2017) Pleistocene climatic changes drive diversification across a tropical savanna. *Molecular Ecology*. In Press. doi: 10.1111/mec.14441.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945 LP-959.

Reich D, Patterson N, Kircher M et al. (2011) Denisova admixture and the first modern human dispersals into southeast asia and oceania. *The American Journal of Human Genetics*, **89**, 516–528.

Rieseberg L (1995) The role of hybridization in evolution: old wine in new skins. *American Journal of Botany*, 82, 944–953.

Ronquist F, Teslenko M, van der Mark P et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Roux C, Fraïsse C, Romiguier J et al. (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, **14**, 1–22.

Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, **11**, 619–625.

Sayyari E, Mirarab S (2016) Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution*, **33**, 1654–1668.

Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.

Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.

Shapiro BJ, Leducq JB, Mallet J (2016) What Is Speciation*? PLoS Genetics*, **12**, 14.

Singhal S (2013) De novo transcriptomic analyses for non-model organisms: An evaluation of methods across a multi-species data set. *Molecular Ecology Resources*, **13**, 403–416.

Singhal S, Bi K (2017) History cleans up messes: The impact of time in driving divergence and introgression in a tropical suture zone. *Evolution*, **71**, 1888–1899.

Singhal S, Moritz C (2013) Reproductive isolation between phylogeographic lineages scales with divergence. *Proceedings of the Royal Society of London Series B*, Biological Sciences, **280**, 20132246.

Slater GSC, Birney E, Box G et al. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution*, **64**, 295–315.

Solís-Lemus C, Ané C (2016) Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting. *PLoS Genetics*, **12**, e1005896.

Solís-Lemus C, Bastide P, Ané C (2017) PhyloNetworks: A package for phylogenetic networks. *Molecular Biology and Evolution*, **34**, 3292–3298.

Solís-Lemus C, Yang M, Ané C (2016) Inconsistency of species tree methods under gene flow. *Systematic Biology*, **65**, 843–851.

Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404–414.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–3.

Stenz NWM, Larget B, Baum D a., Ané C (2015) Exploring tree-like and non-tree-like patterns using genome sequences: an example using the inbreeding plant species *Arabidopsis thaliana* (L.) Heynh. *Systematic Biology*, **64**, 809–823.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510–524.

Than C, Ruths D, Nakhleh L (2008) PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*, **9**, 322.

The *Heliconius* Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94 - 98.

Tung J, Barreiro LB (2017) The contribution of admixture to primate evolution. *Current Opinion in Genetics and Development*, **47**, 61–68.

Ward BJ, van Oosterhout C (2016) hybridcheck: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Molecular Ecology Resources*, **16**, 534–539.

Wen D, Nakhleh L (2017) Coestimating Reticulate Phylogenies and Gene Trees from Multilocus Sequence Data. *Systematic Biology*, syx085.

Wen D, Yu Y, Nakhleh L (2016) Bayesian inference of reticulate phylogenies under the multispecies network coalescent. *PLoS Genetics*, **12**, 1–17.

Wickham H (2017) tidyverse: easily install and load the "tidyverse" packages. *R package version 1.2.1*.

Wilson S, Swan G (2017) *A complete guide to reptiles of Australia*. New Holland.

Yu Y, Dong J, Liu KJ, Nakhleh L (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences*, **111**, 16448–16453.

Yu Y, Nakhleh L (2015) A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics*, **16**, S10.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, **18**, 821–9.

Zhang C, Ogilvie HA, Drummond AJ, Stadler T (2017) Bayesian inference of species networks from multilocus sequence data. *Molecular Biology and Evolution*, msx307.

Acknowledgements

We are grateful to members of the Moritz lab for ongoing discussions relating to this study, and in particular to Joshua Peñalba, Huw Ogilvie and Rebecca Laver for advice on analysis or comments on the manuscript. A.C.A.S is supported by the FCT grant SFRH/BD/88740/2012 and this research was supported by an ARC Laureate Fellowship awarded to C.M. (ARC FL110100104). This work used the ACRF Biomolecular Resource Facility of the Australian National University.

Data Accessibility

Raw sequencing reads associated with this study are available at NCBI Sequence Read Archive (http:// www.ncbi.nlm.nih.gov/sra; Project Accession: BioProject PRJNA289283; BioSamples SAMN04420533 - SAMN08007763). Exon capture processing code and different datasets and scripts that were used in data analysis are available from the Dryad Digital Repository: http://dx.doi.org/XXXXX (waiting approval of manuscript).

Author Contributions

A.C.A.S. designed the research, did laboratorial work, data analyses, and wrote the manuscript. S.P. designed the research, did laboratorial work, advised in data analyses and commented on the manuscript. M.B. designed the research, did data analyses and commented on the manuscript. J.G.B. did the initial bioinformatics and commented on the manuscript. M.M.C. and C.F. had intellectual input into the direction of this study and commented on the manuscript. C.M. designed the research, assisted in writing the manuscript and funded the work.

.

5.6 Supplementary Material

Appendix S5.1 Information on the samples used in this study, including NCBI SRA Accession, sources of published data, tissue voucher number, mean sequence coverage, heterozygosity, number of loci, geographic location, species and lineage ID, and analyses in which they were used (AB – ABBA-BABA; A – ASTRAL; S – SplitsTree, Structure and PCoA; P – PhyloNetworks). B18 – Bragg et al. 2018, P16 – Potter et al. 2016, P17 – Potter et al. 2017, AS17 – Afonso Silva et al. 2017.

Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
SP08_16	SAMN05207344	B18	ABTC11021	250.95	0.0025	2402	-16.8167	145.6333	*rubrigularis*	outgroup		A/P/AB
SP02B_25	SAMN08007692		ABTC29892	111	0.0033	2909	-11.9019	136.4536	*amax*	AECI		S
SP04_11	SAMN04420549	P16	ABTC29886	192.18	0.0018	2939	-11.7769	136.5342	*amax*	AECI		S
SP04_17	SAMN04420551	P16	ABTC29931	173.72	0.0022	2939	-11.8769	136.4222	*amax*	AECI		S/A
SP03_29	SAMN04420535	P17	ABTC11835	81.73	0.0078	2750	-12.7	132.9333	*amax*	AETE		S
SP03_34	SAMN04420570	P17	R27578	125.71	0.0037	2918	-12.72	131.7022	*amax*	AETE	At	S/AB
SP03_35	SAMN04420552	P17	ABTC30120	183.87	0.0045	2925	-14.8	134.945	*amax*	AETE		S
SP04_1	SAMN04420559	P17	ABTC72515	114.03	0.0052	2924	-13.55	133.3667	*amax*	AETE		S
SP04_13	SAMN04420556	P17	ABTC30668	99.5	0.0054	2887	-12.6961	134.8211	*amax*	AETE		S
SP04_16	SAMN04420543	P17	ABTC28521	130.79	0.0044	2929	-12.1731	134.5844	*amax*	AETE		S
SP04_22	SAMN04420555	P17	ABTC30661	147.1	0.005	2927	-12.3697	134.9164	*amax*	AETE	At	S/AB
SP04_23	SAMN04420539	P17	ABTC21782	157.15	0.0025	2933	-12.3442	136.978	*amax*	AETE		S/A/P
SP04_3	SAMN04420541	P17	ABTC28331	128.39	0.005	2924	-13.7886	132.8529	*amax*	AETE	At	S/AB
SP04_32	SAMN04420553	P17	ABTC30200	110.45	0.005	2918	-12.3825	133.0758	*amax*	AETE		S
SP04_44	SAMN04420540	P17	ABTC28225	91.71	0.0047	2871	-11.57	132.87	*amax*	AETE		S
SP04_49	SAMN04420579	P17	R36835	79.52	0.007	2836	-13.9656	131.6951	*amax*	AETE	At	S/AB
SP04_52	SAMN04420534	P17	A003410	166.54	0.0086	2919	-12.5814	134.3081	*amax*	AETE		S
SP04_55	SAMN04420533	P16	A002942	94.2	0.004	2902	-12.05	134.2167	*amax*	AETE		S
SP04_6	SAMN04420546	P17	ABTC29173	164.39	0.005	2929	-12.3256	133.0541	*amax*	AETE		S
SP11_29	SAMN05207399	B18	ABTC72498	181.49	0.0057	2878	-13.3833	133.4167	*amax*	AETE		S
SP11_30	SAMN05207400	B18	CCM1996	158.44	0.005	2899	-12.8877	132.9179	*amax*	AETE		S
SP11_31	SAMN07961671	P17	CCM2362	119.32	0.0049	2868	-12.9947	132.8836	*amax*	AETE	At	S/AB
SP11_32	SAMN07961672	P17	CCM2385	83.4	0.0053	2843	-12.8442	132.9877	*amax*	AETE		S
SP11_53	SAMN07961693	P17	CCM2347	94.35	0.0055	2849	-13.0955	132.9943	*amax*	AETE		S
SP11_54	SAMN07961694	P17	CCM2345	185.17	0.0054	2911	-13.0955	132.9943	*amax*	AETE		S
SP03_26	SAMN04420538	P16	ABTC11961	127.79	0.0083	2886	-17.4667	138.3333	*amax*	AGULF		S
SP03_27	SAMN04420536	P17	ABTC11897	133.42	0.0063	2897	-16.6667	135.85	*amax*	AGULF		S

Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
SP03_28	SAMN04420537	P17	ABTC11900	86.01	0.006	2857	-16.45	134.6167	*amax*	AGULF		S
SP03_30	SAMN04420560	P16	ABTC72760	225.7	0.0027	2909	-20.775	139.7856	*amax*	AGULF		S
SP03_40	SAMN04420573	P17	R29908	107.1	0.0053	2845	-15.8498	135.293	*amax*	AGULF		S
SP04_18	SAMN04420544	P16	ABTC28693	219.27	0.0042	2931	-15.53	135.42	*amax*	AGULF		S/A
SP04_21	SAMN08007693		ABTC29574	259.85	0.0046	2116	-18.6	138.13	*amax*	AGULF		S
AS02_27	SAMN08007694		CCM0999	140.51	0.0038	2917	-14.8856	126.3599	*amax*	AWTEK		S
AS02_31	SAMN08007695		CCM1064	211.74	0.0037	2891	-15.2609	126.201	*amax*	AWTEK	Ak	S/AB
AS02_32	SAMN08007696		CCM1069	165.08	0.0042	2896	-15.8512	127.4103	*amax*	AWTEK		S
AS02_33	SAMN08007697		CCM1129	208.84	0.0038	2883	-14.7811	126.6349	*amax*	AWTEK	Ak	S/AB
AS02_35	SAMN08007698		CCM1188	121.19	0.005	2904	-14.8317	125.7192	*amax*	AWTEK		S
AS02_46	SAMN08007699		CCM1406	307.55	0.0044	2838	-16.496	125.3393	*amax*	AWTEK	Ak	S/AB
AS02_48	SAMN08007700		CCM1642	283.8	0.0042	2867	-16.1711	125.9876	*amax*	AWTEK		S
SP02B_38	SAMN08007701		R36558	159.78	0.0052	2859	-15.6577	129.6593	*amax*	AWTEK		S
SP03_32	SAMN04420577	P17	R36595	121.53	0.0042	2917	-15.8752	129.051	*amax*	AWTEK	Ak	S/AB
SP03_36	SAMN04420576	P16	R36590	139.33	0.0048	2900	-15.6119	131.1161	*amax*	AWTEK		S
SP03_48	SAMN08007702		CMWA33	55.27	0.0099	1960	-16.2518	126.4908	*amax*	AWTEK		S
SP04_10	SAMN04420547	P17	ABTC29644	178.48	0.0042	2936	-13.4813	130.6987	*amax*	AWTEK		S
SP04_12	SAMN04420558	P17	ABTC67991	279.27	0.0046	2720	-15.3489	130.2744	*amax*	AWTEK		S
SP04_14	SAMN04420548	P17	ABTC29705	184.44	0.0045	2926	-13.4086	130.8967	*amax*	AWTEK		S
SP04_15	SAMN04420542	P17	ABTC28403	180.39	0.0036	2940	-14.7877	125.8282	*amax*	AWTEK		S/A/P
SP04_27	SAMN04420566	P17	R146055	92.8	0.0044	2891	-16.7472	123.7956	*amax*	AWTEK		S
SP04_29	SAMN04420567	P17	R166926	134.52	0.0036	2937	-14.8219	125.7219	*amax*	AWTEK		S
SP04_31	SAMN04420564	P17	R132767	105.3	0.0031	2927	-15.2908	128.6828	*amax*	AWTEK		S
SP04_33	SAMN04420563	P16	R114401	119.92	0.0043	2909	-16.15	123.75	*amax*	AWTEK	Ak	S/AB
SP04_34	SAMN04420569	P16	R171890	141.29	0.0025	2918	-16.0894	124.0931	*amax*	AWTEK		S
SP04_37	SAMN04420571	P16	R28957	123.35	0.004	2780	-12.7152	130.856	*amax*	AWTEK		S
SP04_48	SAMN08007703		R29345	136.99	0.0038	2489	-12.8677	131.5847	*amax*	AWTEK		S
SP05_1	SAMN04420545	P17	ABTC29098	40.08	0.0037	2634	-11.85	132.8	*gracilis*	GETE		S
SP05_2	SAMN04420554	P16	ABTC30644	123.78	0.002	2518	-12.3706	134.9169	*gracilis*	GETE		S
SP05_4	SAMN07961643	P17	ABTC29165	39.18	0.004	2518	-12.55	132.916	*gracilis*	GETE		S

Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
SP11_1	SAMN05207392	B18	ABTC11793	139.53	0.003	2916	-12.7	132.9333	*gracilis*	GETE		S/A
SP11_2	SAMN07961660	P17	ABTC28954	215.38	0.0026	2897	-11.25	132.22	*gracilis*	GETE		S
SP11_3	SAMN07961669	P17	ABTC28507	100.63	0.0029	2849	-12.12	133.68	*gracilis*	GETE		S
SP11_4	SAMN07961679	P17	ABTC30715	167.87	0.0024	2911	-12.2433	134.9856	*gracilis*	GETE		S
AS01_52	SAMN08007704		CCM0961	45.49	0.0051	2859	-14.7389	126.4666	*gracilis*	GKIM		S*
AS02_11	SAMN08007705		CCM0786	181.61	0.0009	2928	-15.2237	126.2182	*gracilis*	GKIM		S
AS02_13	SAMN08007706		CCM0794	110.98	0.0012	2924	-15.1174	126.1297	*gracilis*	GKIM		S
AS02_14	SAMN08007707		CCM0798	97.12	0.0012	2926	-15.1328	126.1478	*gracilis*	GKIM		S
AS02_23	SAMN08007708		CCM0955	153.97	0.0009	2924	-14.7387	126.4689	*gracilis*	GKIM		S
AS02_24	SAMN08007709		CCM0959	147.28	0.001	2933	-14.7394	126.4706	*gracilis*	GKIM		S
AS02_26	SAMN08007710		CCM0969	139.09	0.0009	2927	-14.7423	126.4669	*gracilis*	GKIM	Gk	S/AB
AS02_3	SAMN08007711		CCM0734	156.13	0.0012	2929	-15.2024	125.9019	*gracilis*	GKIM		S
AS02_4	SAMN08007712		CCM0740	216.75	0.001	2925	-15.1968	125.9017	*gracilis*	GKIM		S
AS02_5	SAMN08007713		CCM0746	324.55	0.0009	2909	-15.1174	126.1297	*gracilis*	GKIM		S
AS02_8	SAMN08007714		CCM0778	194.1	0.001	2933	-15.1994	125.9032	*gracilis*	GKIM		S
SP05_11	SAMN07961616	P17	R36713	31.95	0.0047	2435	-14.0739	130.7859	*gracilis*	GKIM		S
SP05_23	SAMN05207316	B18	CCM1135	44.68	0.0016	2800	-14.7813	126.6358	*gracilis*	GKIM	Gk	S/AB
SP05_7	SAMN07961647	P17	R117670	54.38	0.0011	2764	-14.8072	125.8417	*gracilis*	GKIM	Gk	S/AB
SP05_9	SAMN05207317	B18	CCM0957	103.74	0.001	2941	-14.7394	126.4706	*gracilis*	GKIM		S/A/P
SP11_5	SAMN07961689	P17	CCM0735	172.74	0.0011	2908	-15.2024	125.9019	*gracilis*	GKIM	Gk	S/AB
SP11_51	SAMN07961691	P17	R36706	86.54	0.0013	2927	-14.2563	130.8934	*gracilis*	GKIM	Gk	S/AB
SP11_55	SAMN08007715		CCM0748	127.66	0.0008	2918	-15.1283	126.1393	*gracilis*	GKIM		S
SP11_6	SAMN05207406	B18	CCM0958	119.55	0.0007	2915	-14.7394	126.4706	*gracilis*	GKIM		S
SP11_7	SAMN07961697	P17	CCM1134	117.43	0.0008	2909	-14.7813	126.6358	*gracilis*	GKIM		S
SP05_13	SAMN07961618	P17	ABTC30074	118.88	0.0038	2927	-11.7822	130.7744	*gracilis*	GMEL		S/A
SP11_13	SAMN07961653	P17	ABTC29973	87.31	0.0037	2152	-11.595	130.6719	*gracilis*	GMEL		S
SP11_14	SAMN07961654	P17	ABTC30091	123.05	0.003	2910	-11.4178	130.5117	*gracilis*	GMEL		S
SP11_15	SAMN07961655	P17	ABTC30079	139.06	0.0031	2887	-11.5442	130.5875	*gracilis*	GMEL		S
SP05_12	SAMN07961617	P17	CCM0522	122.84	0.0037	2914	-12.3697	130.8725	*gracilis*	GWTE	Gt	S/AB
SP05_14	SAMN05207314	B18	CCM0457	110.08	0.0028	2940	-13.4352	132.4143	*gracilis*	GWTE	Gt	S/A/P/AB

Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
SP11_10	SAMN05207393	B18	ABTC28428	166.48	0.0032	2912	-13.2	132.05	*gracilis*	GWTE		S
SP11_11	SAMN07961651	P17	ABTC29615	134.44	0.0029	2852	-13.4813	130.6987	*gracilis*	GWTE	Gt	S/AB
SP11_12	SAMN07961652	P17	ABTC81197	143.32	0.0032	2892	-12.3051	132.067	*gracilis*	GWTE		S
SP11_8	SAMN07961698	P17	ABTC28276	90.23	0.0034	2814	-12.58	131.83	*gracilis*	GWTE	Gt	S/AB
SP11_9	SAMN07961699	P17	CCM0512	170.71	0.0031	2927	-12.9003	131.6757	*gracilis*	GWTE	Gt	S/AB
AS01_42	SAMN06927849	AS17	R158583	34.91	0.0055	2617	-14.2794	125.3061	*johnstonei (s. l.)*	insularis	Ji	S/AB
AS01_44	SAMN06927851	AS17	R158705	27.26	0.0042	2171	-14.3983	124.9775	*johnstonei (s. l.)*	insularis	Ji	S/AB
AS01_45	SAMN06927852	AS17	R117953	87.72	0.0006	2925	-14.1858	125.7347	*johnstonei (s. l.)*	insularis	Ji	S/P/AB
AS01_46	SAMN06927853	AS17	R117967	89.91	0.0014	2944	-14.1855	125.7244	*johnstonei (s. l.)*	insularis	Ji	S/A/P/AB
AS01_18	SAMN06927825	AS17	R117787	30.35	0.0044	2494	-14.3802	125.9511	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_19	SAMN06927826	AS17	R158783	40.82	0.0041	2854	-14.6019	125.2039	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_20	SAMN06927827	AS17	BP00664	49.35	0.0028	2920	-14.9763	124.9131	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_22	SAMN06927829	AS17	R171227	31.54	0.0037	2558	-15.3513	124.5269	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_25	SAMN06927832	AS17	CCM0737	50.38	0.0016	2906	-15.1989	125.9035	*johnstonei (s. l.)*	johnstonei *(s. s.)*	Jk	S/AB
AS01_27	SAMN06927834	AS17	CCM1125	62.49	0.003	2876	-14.7806	126.6369	*johnstonei (s. l.)*	johnstonei *(s. s.)*	Jk	S/AB
AS01_28	SAMN06927835	AS17	CCM1196	47.76	0.0048	2783	-14.8211	125.7214	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_29	SAMN06927836	AS17	R171237	62.85	0.0025	2929	-15.2597	124.4453	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_30	SAMN06927837	AS17	CCM0934	92.25	0.0039	2932	-14.5221	126.4612	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_31	SAMN06927838	AS17	R117726	91.23	0.002	2933	-14.1638	125.6797	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_32	SAMN06927839	AS17	R117838	33.11	0.0034	2544	-14.3755	125.9897	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_47	SAMN06927854	AS17	R173481	79.98	0.0027	2942	-14.6082	126.9317	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S/A/P
AS01_48	SAMN06927855	AS17	R173484	27.21	0.0034	2037	-14.6082	126.9317	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_51	SAMN06927857	AS17	BP01423	46	0.0047	2863	-15.9497	124.5611	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS01_56	SAMN06927858	AS17	R169980	34.28	0.0051	2539	-16.1766	123.6394	*johnstonei (s. l.)*	johnstonei *(s. s.)*	Jk	S*/AB
AS02_36	SAMN08007716		CCM1197	252.47	0.0045	2902	-14.8211	125.7211	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS02_38	SAMN08007717		CCM1263	188.05	0.0024	2910	-17.0684	125.2462	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS02_39	SAMN08007718		CCM1264	292.44	0.0025	2880	-17.0684	125.2462	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS02_40	SAMN08007719		CCM1265	262.04	0.0022	2905	-17.0684	125.2462	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS02_41	SAMN08007720		CCM1267	247.42	0.0024	2917	-17.0697	125.2461	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS02_52	SAMN08007721		CCM5837	201.82	0.0029	2914	-14.761	126.5271	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S

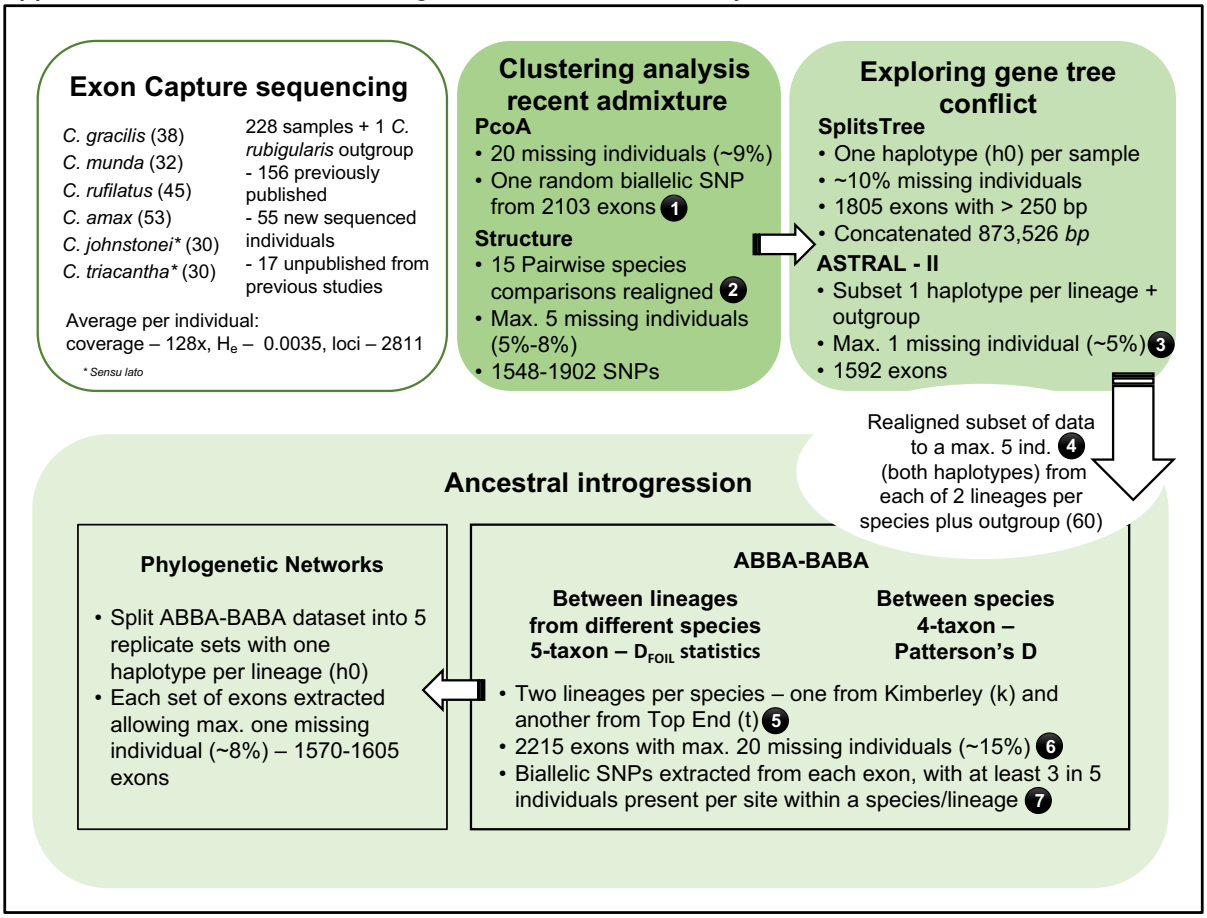
Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
AS02_53	SAMN08007722		CCM5857	124.09	0.0031	2930	-14.7767	126.5139	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
SP03_1	SAMN08007723		R117843	79.81	0.0056	2714	-14.3491	125.9592	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
SP05_8	SAMN08007724		CCM1243	45.15	0.0027	2809	-16.9982	125.2067	*johnstonei (s. l.)*	johnstonei *(s. s.)*	Jk	S/AB
SP16_41	SAMN08007725		BP02304	159.7	0.0022	2921	-14.4594	127.8597	*johnstonei (s. l.)*	johnstonei *(s. s.)*	Jk	S/AB
SP16_42	SAMN08007726		BP02482	174.41	0.0015	2923	-14.4852	127.8218	*johnstonei (s. l.)*	johnstonei *(s. s.)*		S
AS02_19	SAMN08007727		CCM0915	180.35	0.0023	2908	-14.8127	126.3288	*munda*	MBROAD	Mk	S/AB
AS02_42	SAMN08007728		CCM1355	232.56	0.0027	2905	-16.4893	125.3528	*munda*	MBROAD		S
AS02_50	SAMN08007729		R168497	127.63	0.0036	2914	-17.4277	122.1522	*munda*	MBROAD		S/A/P
AS02_54	SAMN08007730		CCM5899	175.23	0.0021	2909	-14.8133	126.326	*munda*	MBROAD		S
AS02_56	SAMN08007731		CCM5902	225.7	0.0022	2903	-14.8162	126.3272	*munda*	MBROAD		S
SP04_43	SAMN05207312	B18	R131750	70.21	0.0024	2875	-22.173	117.7225	*munda*	MBROAD		S
SP04_51	SAMN08007732		R35682	76.65	0.0087	2886	-14.7718	134.8052	*munda*	MBROAD		S*
SP05_10	SAMN07961615	P17	CCM1354	29.05	0.005	2284	-16.4942	125.3417	*munda*	MBROAD	Mk	S/AB
SP05_32	SAMN07961635	P17	R145677	37.64	0.0032	2750	-22.3	119.0333	*munda*	MBROAD		S
SP05_33	SAMN07961636	P17	ABTC77098	66.55	0.0041	2400	-18.7286	144.3122	*munda*	MBROAD		S
SP07_10	SAMN05207318	B18	ABTC10845	76.91	0.0035	2390	-20.533	145.4	*munda*	MBROAD		S
SP08_10	SAMN05207338	B18	ABTC10846	66.08	0.004	2380	-20.533	145.4	*munda*	MBROAD		S
SP11_33	SAMN05207401	B18	ABTC102927	105.79	0.0036	2641	-19.229	140.3482	*munda*	MBROAD		S
SP11_34	SAMN07961674	P17	ABTC76982	86.85	0.0039	2904	-21.5	145.0342	*munda*	MBROAD		S
SP11_35	SAMN07961675	P17	R157468	32.06	0.006	2455	-19.5891	128.8603	*munda*	MBROAD	Mk	S*/AB
SP11_37	SAMN07961676	P17	ABTC28871	96.38	0.0039	2805	-15.62	136.35	*munda*	MBROAD		S
SP11_38	SAMN07961677	P17	A007290	133.96	0.004	2888	-15.4611	142.1756	*munda*	MBROAD		S
SP11_39	SAMN07961678	P17	ABTC28037	97.31	0.0021	2878	-15	129.5833	*munda*	MBROAD	Mk	S/AB
SP11_40	SAMN07961680	P17	R114219	52.49	0.0039	2857	-16.4	122.9333	*munda*	MBROAD	Mk	S/AB
SP05_38	SAMN07961641	P17	ABTC28218	54.25	0.0051	2880	-11.57	132.87	*munda*	MTE		S
SP05_39	SAMN07961642	P17	ABTC30656	87.23	0.0039	2920	-12.3706	134.9206	*munda*	MTE	Mt	S/AB
SP05_40	SAMN07961644	P17	CCM0515	175.44	0.0048	2926	-12.9046	131.6513	*munda*	MTE	Mt	S/A/P/AB
SP05_41	SAMN07961645	P17	ABTC28447	26.28	0.0057	1941	-12.89	135.47	*munda*	MTE		S
SP11_44	SAMN05207402	B18	ABTC72508	143.76	0.0051	2916	-13.6333	133.0333	*munda*	MTE	Mt	S/AB
SP11_45	SAMN05207403	B18	ABTC28277	151.27	0.0048	2896	-12.58	131.83	*munda*	MTE	Mt	S/AB

Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
SP11_46	SAMN05207404	B18	ABTC10977	182.98	0.0035	2911	-12.55	132.916	*munda*	MTE	Mt	S/AB
SP11_47	SAMN05207405	B18	ABTC28960	137.61	0.004	2888	-11.15	132.15	*munda*	MTE		S
SP11_50	SAMN07961690	P17	ABTC30640	28.93	0.0031	2141	-13.9833	132.7167	*munda*	MTE		S
SP05_36	SAMN07961639	P17	ABTC30095	44.45	0.0043	2848	-11.5333	130.5833	*munda*	MMEL		S
SP05_37	SAMN07961640	P17	ABTC29957	58.23	0.0043	2883	-11.6	130.7	*munda*	MMEL		S
SP11_48	SAMN07961687	P17	ABTC30100	84.56	0.0042	2907	-11.5442	130.5875	*munda*	MMEL		S/A
SP11_49	SAMN07961688	P17	ABTC30035	123.77	0.0043	2903	-11.8294	130.9086	*munda*	MMEL		S
SP05_29	SAMN07961631	P17	ABTC29243	91.94	0.0046	2922	-11.28	136.59	*rufilatus*	RECI		S/A
SP05_30	SAMN07961633	P17	ABTC29275	51.42	0.0035	2783	-11.5	136.44	*rufilatus*	RECI		S
SP05_31	SAMN07961634	P17	ABTC29330	140.62	0.0032	2918	-11.65	136.18	*rufilatus*	RECI		S
SP11_16	SAMN05207394	B18	ABTC29321	170.41	0.0027	2891	-11.627	136.1905	*rufilatus*	RECI		S
SP11_17	SAMN05207395	B18	ABTC29295	135.76	0.0029	2901	-11.65	136.18	*rufilatus*	RECI		S
SP11_18	SAMN07961658	P17	ABTC29271	188.52	0.0024	2905	-11.2647	136.6223	*rufilatus*	RECI		S
SP11_19	SAMN07961659	P17	ABTC29114	77.61	0.004	2862	-11.2647	136.6223	*rufilatus*	RECI		S
SP04_30	SAMN08007691	P17	R35296	152.17	0.0027	2938	-12.7	130.9833	*rufilatus*	RTE		S/A/P
SP04_40	SAMN07961612	P17	ABTC29122	298.7	0.003	2520	-12.7624	130.5251	*rufilatus*	RTE	Rt	S/AB
SP05_15	SAMN07961620	P17	R35002	54	0.0025	2929	-14.0033	131.2388	*rufilatus*	RTE	Rt	S/AB
SP05_16	SAMN05207315	B18	R27658	151.17	0.0028	2926	-12.523	131.267	*rufilatus*	RTE	Rt	S/AB
SP05_17	SAMN07961622	P17	ABTC29941	139.06	0.0075	2909	-11.6053	130.7119	*rufilatus*	RTE		S
SP11_23	SAMN07961663	P17	ABTC29112	121.42	0.0029	2912	-12.45	130.83	*rufilatus*	RTE	Rt	S/AB
SP11_25	SAMN07961665	P17	ABTC29944	129.24	0.0022	2908	-11.6053	130.7119	*rufilatus*	RTE		S
SP11_26	SAMN07961666	P17	R35650	107.66	0.0024	2894	-12.2018	131.3267	*rufilatus*	RTE		S
SP11_27	SAMN07961667	P17	ABTC29479	124.53	0.0026	2878	-13.4813	130.6987	*rufilatus*	RTE	Rt	S/AB
AS01_53	SAMN08007733		CCM0979	51.23	0.0031	2895	-14.7519	126.4772	*rufilatus*	RKIM		S
AS02_1	SAMN08007734		CCM0671	285.39	0.0019	2887	-15.8748	129.0513	*rufilatus*	RKIM		S
AS02_15	SAMN08007735		CCM0812	238.2	0.0021	2905	-15.1276	126.1352	*rufilatus*	RKIM	Rk	S/AB
AS02_16	SAMN08007736		CCM0899	261.11	0.0021	2921	-14.8522	126.2785	*rufilatus*	RKIM		S
AS02_17	SAMN08007737		CCM0900	138.06	0.0027	2929	-14.8522	126.2785	*rufilatus*	RKIM		S
AS02_18	SAMN08007738		CCM0901	186.58	0.0022	2924	-14.8522	126.2785	*rufilatus*	RKIM		S
AS02_2	SAMN08007739		CCM0673	281.61	0.002	2900	-15.8748	129.0513	*rufilatus*	RKIM		S

Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
AS02_20	SAMN08007740		CCM0923	161.51	0.0023	2924	-14.5209	126.4613	*rufilatus*	RKIM		S
AS02_21	SAMN08007741		CCM0928	245.82	0.0023	2918	-14.7396	126.4677	*rufilatus*	RKIM		S
AS02_22	SAMN08007742		CCM0935	181.98	0.0022	2903	-14.5209	126.4613	*rufilatus*	RKIM		S
AS02_34	SAMN08007743		CCM1131	289.93	0.0021	2912	-14.7806	126.6369	*rufilatus*	RKIM	Rk	S/AB
AS02_49	SAMN08007744		CCM1847	141.37	0.0019	2920	-15.3442	126.1017	*rufilatus*	RKIM		S
AS02_51	SAMN08007745		CCM5768	165.6	0.0022	2909	-14.7924	126.4482	*rufilatus*	RKIM		S
AS02_55	SAMN08007746		CCM5901	150.98	0.0021	2916	-14.8147	126.3272	*rufilatus*	RKIM		S
AS02_6	SAMN08007747		CCM0752	238.02	0.002	2905	-15.1361	126.1523	*rufilatus*	RKIM		S
AS02_7	SAMN08007748		CCM0760	150.81	0.002	2917	-15.3382	126.5881	*rufilatus*	RKIM		S
AS02_9	SAMN08007749		CCM0780	295.13	0.0019	2896	-15.1993	125.8994	*rufilatus*	RKIM		S
SP04_39	SAMN07961611	P17	ABTC29864	172.44	0.0021	2945	-15.7642	129.0972	*rufilatus*	RKIM	Rk	S/A/P/AB
SP04_9	SAMN05207313	B18	CMWA35	168.03	0.0016	2928	-16.2518	126.4908	*rufilatus*	RKIM		S
SP05_18	SAMN07961623	P17	ABTC29863	108.67	0.0023	2929	-15.7642	129.0972	*rufilatus*	RKIM		S
SP05_24	SAMN07961626	P17	ABTC29856	29.89	0.0028	2389	-15.7642	129.0972	*rufilatus*	RKIM		S
SP05_26	SAMN07961628	P17	R115756	78.25	0.0035	2897	-14.8955	128.6611	*rufilatus*	RKIM		S
SP05_27	SAMN07961629	P17	CCM0846	24.63	0.0024	1954	-15.3882	126.5881	*rufilatus*	RKIM		S
SP11_21	SAMN07961661	P17	CCM0672	50.58	0.0023	2698	-15.8748	129.0513	*rufilatus*	RKIM		S
SP11_22	SAMN05207396	B18	CCM0677	111.56	0.002	2895	-15.8748	129.0513	*rufilatus*	RKIM		S
SP11_28	SAMN05207398	B18	ABTC28405	109.04	0.002	2866	-14.7877	125.8282	*rufilatus*	RKIM	Rk	S/AB
SP11_36	SAMN08007750		CCM0898	93.59	0.0024	2879	-14.8522	126.2785	*rufilatus*	RKIM		S
SP11_42	SAMN08007751		CCM0736	82.11	0.0022	2780	-15.1994	125.9007	*rufilatus*	RKIM		S
SP11_56	SAMN07961695	P17	CCM1524	121.78	0.0013	2924	-16.8312	126.2281	*rufilatus*	RKIM	Rk	S/AB
AS01_10	SAMN06927817	AS17	CCM1643	53.58	0.004	2867	-16.1711	125.9876	*triacantha (s. l.)*	isostriacantha		S
AS01_12	SAMN06927819	AS17	ABTC28797	45.51	0.0039	2872	-18.6	136.1	*triacantha (s. l.)*	isostriacantha		S
AS01_13	SAMN06927820	AS17	ABTC29531	46.8	0.003	2850	-18.6083	137.9883	*triacantha (s. l.)*	isostriacantha		S
AS01_14	SAMN06927821	AS17	R117798	26.81	0.0036	2096	-14.3433	126.0208	*triacantha (s. l.)*	isostriacantha		S
AS01_15	SAMN06927822	AS17	R146022	31.18	0.006	2500	-16.6833	123.8333	*triacantha (s. l.)*	isostriacantha		S
AS01_16	SAMN06927823	AS17	R168590	56.94	0.0024	2921	-14.9833	124.9167	*triacantha (s. l.)*	isostriacantha		S
AS01_17	SAMN06927824	AS17	R168910	40.28	0.0038	2866	-14.6	125.1167	*triacantha (s. l.)*	isostriacantha		S
AS01_2	SAMN06927809	AS17	R158008	42.25	0.0041	2717	-16.1486	123.7794	*triacantha (s. l.)*	isostriacantha	Tk	S/AB

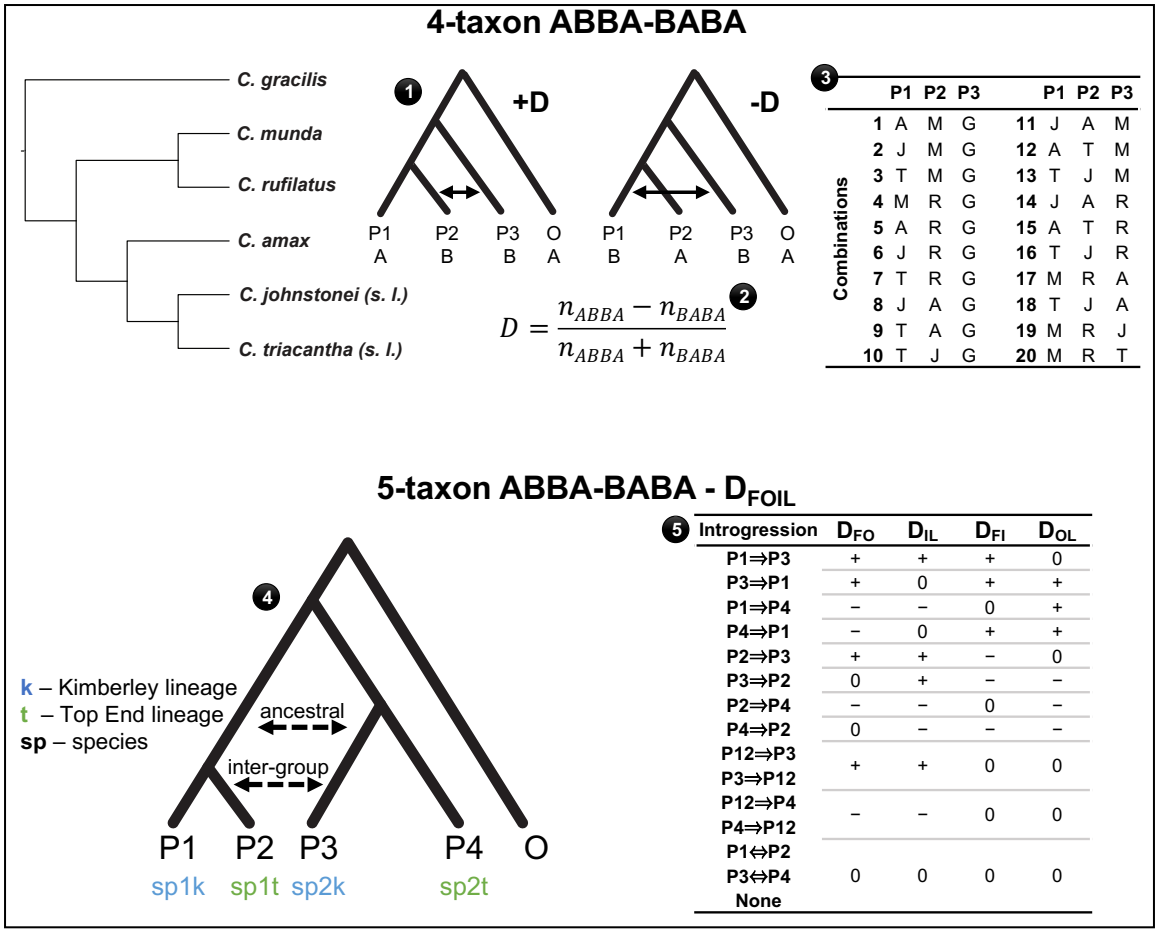
Library	SRA	Source	Sample	Cov	He	Loci	Latitude	Longitude	*Species*	Lineage		Analysis
AS01_5	SAMN06927812	AS17	ABTC29851	55.59	0.0037	2920	-15.9708	129.0403	*triacantha (s. l.)*	isostriacantha	Tk	S/AB
AS01_55	SAMN08007752		CCM1046	29.27	0.0063	2280	-15.1328	126.2002	*triacantha (s. l.)*	isostriacantha		S*
AS01_7	SAMN06927814	AS17	CCM1130	53.87	0.0041	2908	-14.7811	126.6349	*triacantha (s. l.)*	isostriacantha	Tk	S/AB
AS01_8	SAMN06927815	AS17	CCM1202	35.2	0.0042	2661	-14.671	125.7332	*triacantha (s. l.)*	isostriacantha		S
AS01_9	SAMN06927816	AS17	CCM1235	47.2	0.0043	2864	-17.0407	125.2268	*triacantha (s. l.)*	isostriacantha	Tk	S/AB
AS02_10	SAMN08007753		CCM0783	270.89	0.0034	2874	-15.1998	126.0887	*triacantha (s. l.)*	isostriacantha		S
AS02_12	SAMN08007754		CCM0787	171.07	0.0033	2912	-15.1994	125.9032	*triacantha (s. l.)*	isostriacantha	Tk	S/AB
AS02_28	SAMN08007755		CCM1011	145.2	0.0037	2934	-14.88	126.3585	*triacantha (s. l.)*	isostriacantha		S/A/P
AS02_29	SAMN08007756		CCM1020	159.45	0.003	2913	-16.0969	126.5112	*triacantha (s. l.)*	isostriacantha		S
AS02_30	SAMN08007757		CCM1058	173.03	0.0032	2911	-16.0969	126.5112	*triacantha (s. l.)*	isostriacantha		S
AS02_37	SAMN08007758		CCM1201	175.3	0.0037	2917	-14.8317	125.7192	*triacantha (s. l.)*	isostriacantha		S
AS02_43	SAMN08007759		CCM1399	292.04	0.0037	2897	-16.4995	125.3364	*triacantha (s. l.)*	isostriacantha		S
AS02_44	SAMN08007760		CCM1400	315.51	0.0037	2885	-16.4995	125.3364	*triacantha (s. l.)*	isostriacantha		S
AS02_45	SAMN08007761		CCM1401	296.77	0.0035	2880	-16.4995	125.3364	*triacantha (s. l.)*	isostriacantha		S
AS02_47	SAMN08007762		CCM1513	266.97	0.0037	2902	-16.9037	125.7606	*triacantha (s. l.)*	isostriacantha		S
SP04_42	SAMN08007763		BP00182	88.84	0.0046	2781	-16.2564	123.8247	*triacantha (s. l.)*	isostriacantha		S
AS01_34	SAMN06927841	AS17	R139010	44.78	0.0039	2867	-19.8083	121.4639	*triacantha (s. l.)*	triacantha *(s. s.)*	Tt	S/AB
AS01_37	SAMN06927844	AS17	ABTC30066	27.83	0.0031	2241	-11.8522	130.8525	*triacantha (s. l.)*	triacantha *(s. s.)*	Tt	S/AB
AS01_38	SAMN06927845	AS17	ABTC41790	25.96	0.0063	2089	-25.9803	129.4631	*triacantha (s. l.)*	triacantha *(s. s.)*		S*
AS01_39	SAMN06927846	AS17	ABTC29091	49.8	0.0055	2883	-12.6667	132.8833	*triacantha (s. l.)*	triacantha *(s. s.)*	Tt	S*/A/P/AB
AS01_41	SAMN06927848	AS17	CCM1859	40.58	0.0038	2686	-23.5834	134.4994	*triacantha (s. l.)*	triacantha *(s. s.)*	Tt	S/AB
AS01_6	SAMN06927813	AS17	ABTC61613	29.62	0.0023	2358	-15.005	129.5833	*triacantha (s. l.)*	triacantha *(s. s.)*	Tt	S/AB

Appendix S5.2 Flow chart summarizing the data collection and analysis workflow



1. Randomly selected biallelic SNP with the least amount of missing data to a maximum of 20%.
2. Detected that when running STRUCTURE with K = 2, violations of the Hardy-Weinberg equilibrium assumption due to intra-lineage genetic structure were biasing admixture inference (i.e. individuals of the same lineage showing admixed patterns) in the two sets of comparisons between sister species (T-J and M-R). Therefore we removed the individuals belonging to the most divergent lineage in each comparison (*C. insularis* and RECI). To further verify if intraspecific structure could affect admixture inference, pairwise comparisons were also performed with lineages exclusively from either the Kimberley or the Top End (Appendix 5.6 and 5.7).
3. To increase the number of loci above 1000 (0% missing data yielded about 600) while using the most complete loci set.
4. To use an approximately even number of individuals per species, considering that the minimum number of individuals was four (for *C. insularis*).
5. In the D_{FOIL} analyses, all taxa except *C. johnstonei (s. l.)* had a single different lineage distributed in each of the two regions (Kimberley and Top End). We included both *C. johnstonei* (Jk) and *C. insularis* (Ji), which are both from the Kimberley, because *C. johnstonei* broadly overlaps with other species showing some evidence of admixture, and two lineages are needed for the five-taxon ABBA-BABA tests.
6. Initial alignments with all species were subsequently subsetted per pairwise species combination to optimize the extraction of biallelic SNPs for each comparison.
7. Preliminary analyses testing different proportions of missing individuals, by resampling 100x to the same number of SNPs with no missing data (not shown), suggested that SNP sample size has greater effect than the quantity of missing data in the results. Because similar results were obtained with either 20% or 40% missing data per site within each species/lineage, we applied the latter threshold in order to include more SNPs in each comparison.

Appendix S5.3 Schematic diagram depicting the two types of ABBA-BABA analyses used in this study and their results.



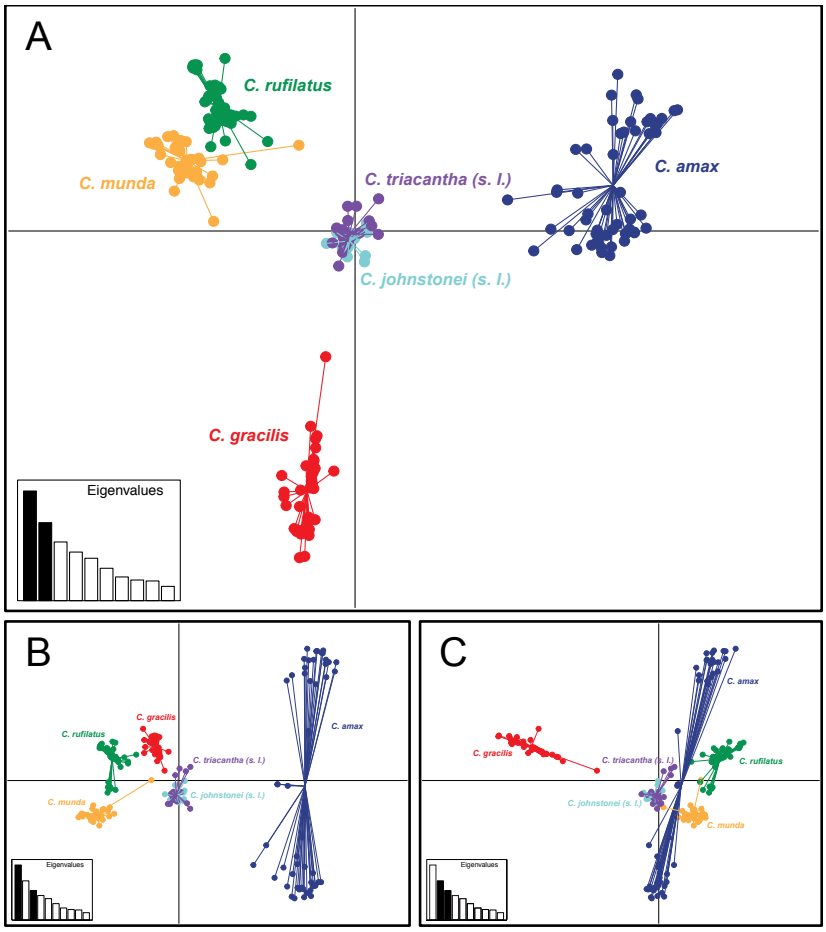
	P1	P2	P3		P1	P2	P3
1	A	M	G	11	J	A	M
2	J	M	G	12	A	T	M
3	T	M	G	13	T	J	M
4	M	R	G	14	J	A	R
5	A	R	G	15	A	T	R
6	J	R	G	16	T	J	R
7	T	R	G	17	M	R	A
8	J	A	G	18	T	J	A
9	T	A	G	19	M	R	J
10	T	J	G	20	M	R	T

Introgression	D_{FO}	D_{IL}	D_{FI}	D_{OL}
P1⇒P3	+	+	+	0
P3⇒P1	+	0	+	+
P1⇒P4	–	–	0	+
P4⇒P1	–	0	+	+
P2⇒P3	+	+	–	0
P3⇒P2	0	+	–	–
P2⇒P4	–	–	0	–
P4⇒P2	0	–	–	–
P12⇒P3 P3⇒P12	+	+	0	0
P12⇒P4 P4⇒P12	–	–	0	0
P1⇔P2 P3⇔P4 None	0	0	0	0

1. The four-taxon ABBA-BABA test requires a pectinate tree to assess introgression between P3 and P2 (ABBA pattern) or P1 (BABA). In the case of discordances due to ILS, it would be expected similar numbers of ABBA and BABA biallelic sites. The test requires a known species tree and does not allow to investigate introgression between sister taxa.
2. Patterson's D has been shown to be the same as the parameter in equation 2 of Durand et al. 2010, in which allele frequencies are used to calculate D, thereby allowing to use more than one individual per species.
3. Table with all possible 20 species combinations based on the known species tree topology.
4. The five-taxon ABBA-BABA test uses a system of four D-like statistics calculated from different patterns in a balanced tree, where P1 and P2 are sister taxa that diverged more recently than P3 and P4 in the sister clade. For each species pairwise comparison, a lineage from the Kimberley (k) and another from the Top End (t) from both species were used to investigate introgression between lineages from different species in the same region.
5. Introgression and its direction are inferred from the combination of results for the four D-like statistics. Ancestral introgressions have ambiguous signals and their direction cannot be ascertained.

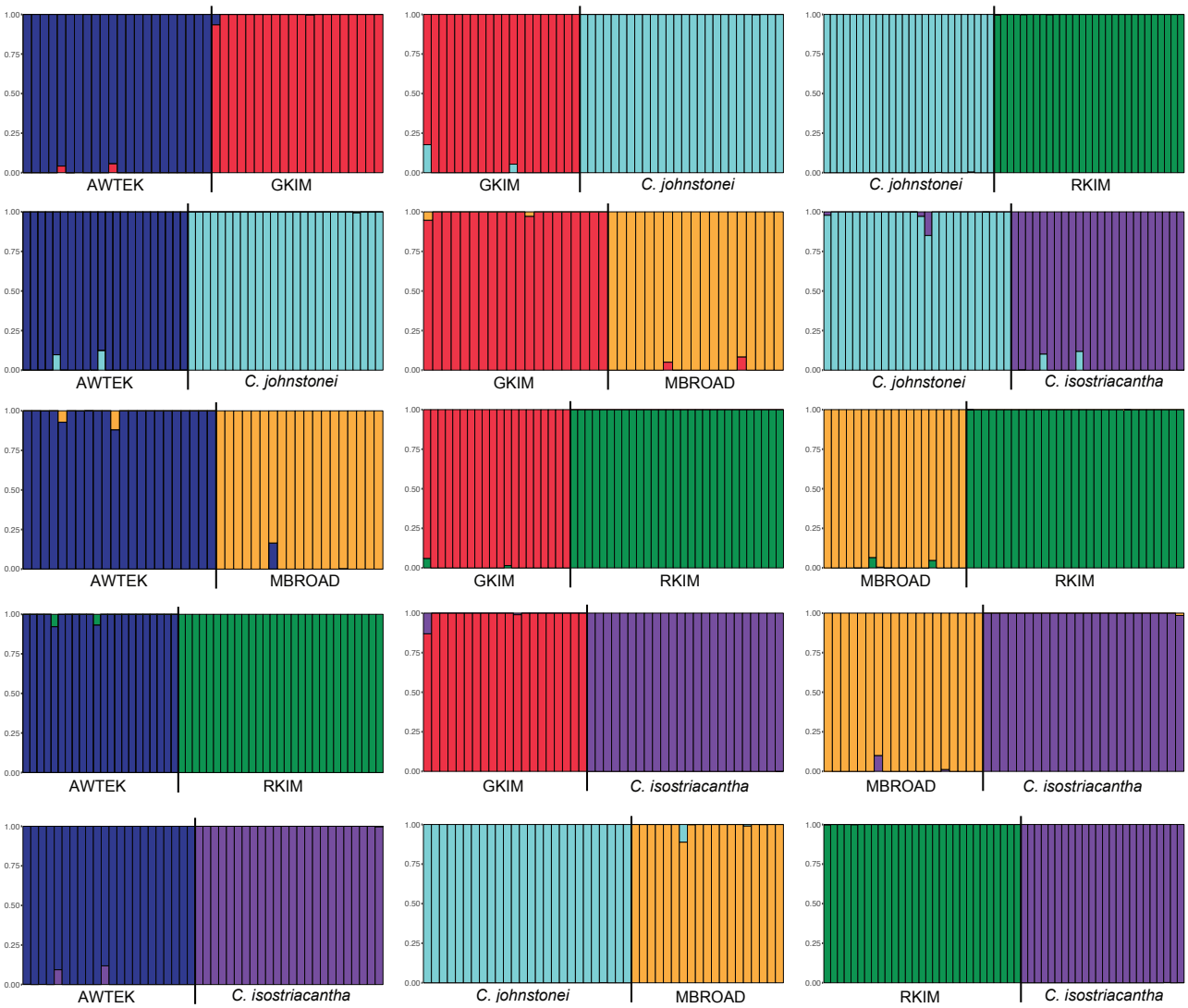
Appendix S5.4 Averages dxy (above diagonal), pi (along diagonal in bold) and da (below diagonal) for the studied species and lineages based on the full dataset and on that used in the ABBA-BABA analysis.

	Ak	At	Gk	Gt	Ji	Jk	Mk	Mt	Rk	Rt	Tt	Tk
Ak	**0.0029**	0.0060	0.0103	0.0103	0.0082	0.0083	0.0093	0.0094	0.0094	0.0094	0.0084	0.0086
At	0.0033	**0.0037**	0.0104	0.0103	0.0082	0.0084	0.0093	0.0094	0.0094	0.0094	0.0084	0.0086
Gk	0.0089	0.0088	**0.0020**	0.0027	0.0099	0.0100	0.0101	0.0104	0.0105	0.0105	0.0102	0.0103
Gt	0.0084	0.0082	0.0015	**0.0026**	0.0098	0.0100	0.0100	0.0104	0.0104	0.0105	0.0101	0.0103
Ji	0.0065	0.0063	0.0091	0.0087	**0.0021**	0.0058	0.0090	0.0091	0.0091	0.0091	0.0062	0.0064
Jk	0.0059	0.0056	0.0084	0.0079	0.0039	**0.0034**	0.0091	0.0092	0.0092	0.0093	0.0063	0.0065
Mk	0.0068	0.0065	0.0084	0.0080	0.0071	0.0064	**0.0033**	0.0038	0.0073	0.0073	0.0093	0.0095
Mt	0.0068	0.0065	0.0087	0.0082	0.0072	0.0065	0.0009	**0.0036**	0.0074	0.0074	0.0094	0.0096
Rk	0.0075	0.0071	0.0094	0.0090	0.0077	0.0071	0.0050	0.0050	**0.0027**	0.0040	0.0094	0.0096
Rt	0.0076	0.0074	0.0096	0.0092	0.0080	0.0073	0.0053	0.0053	0.0023	**0.0021**	0.0095	0.0097
Tt	0.0060	0.0057	0.0087	0.0082	0.0043	0.0037	0.0067	0.0067	0.0073	0.0076	**0.0031**	0.0040
Tk	0.0063	0.0061	0.0091	0.0086	0.0048	0.0041	0.0071	0.0071	0.0077	0.0080	0.0017	**0.0027**

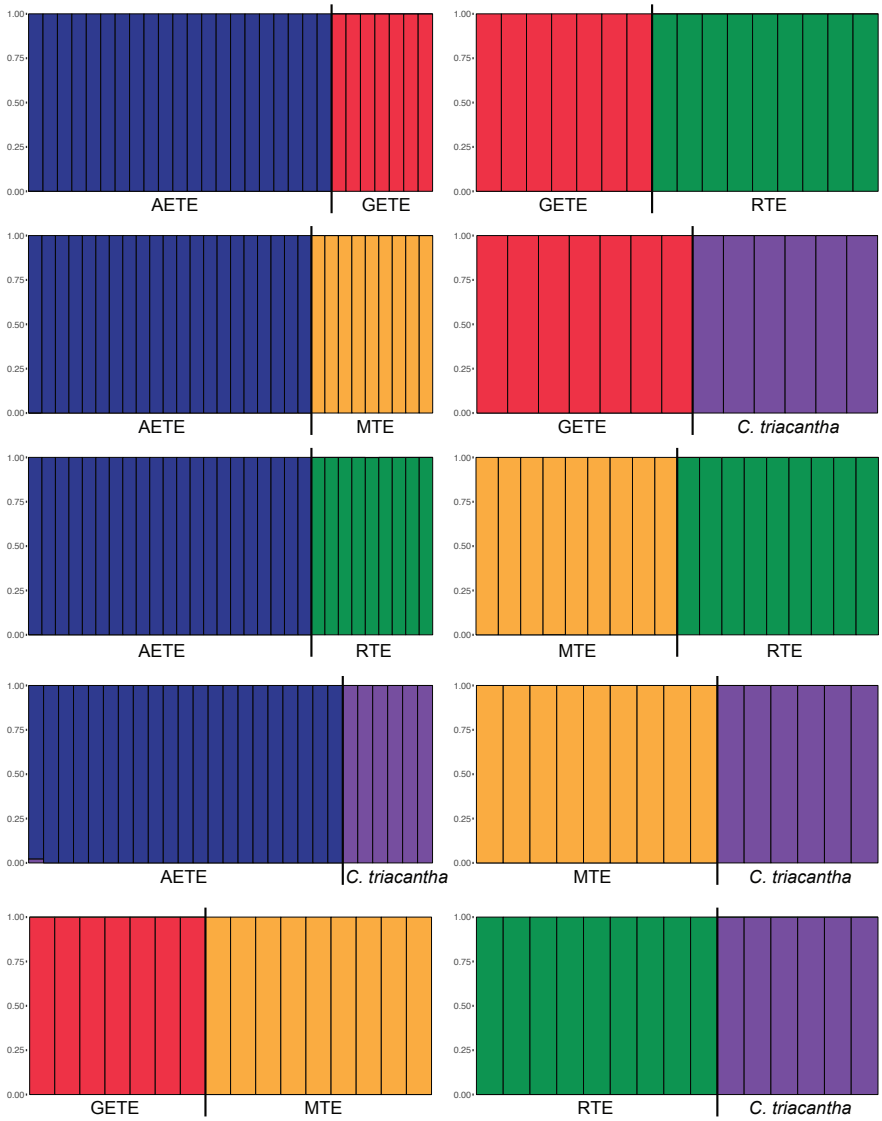


Appendix S5.5 PCoA including all individuals and using one randomly sampled biallelic SNP per locus from 2103 loci. Results are shown for PC 1 and PC 2 (A), PC 1 and PC 3 (B), and PC 2 and PC 3 (C).

Appendix S5.6 Bar plots of STRUCTURE results for pairwise comparisons between species and lineages mainly distributed in the Kimberley.



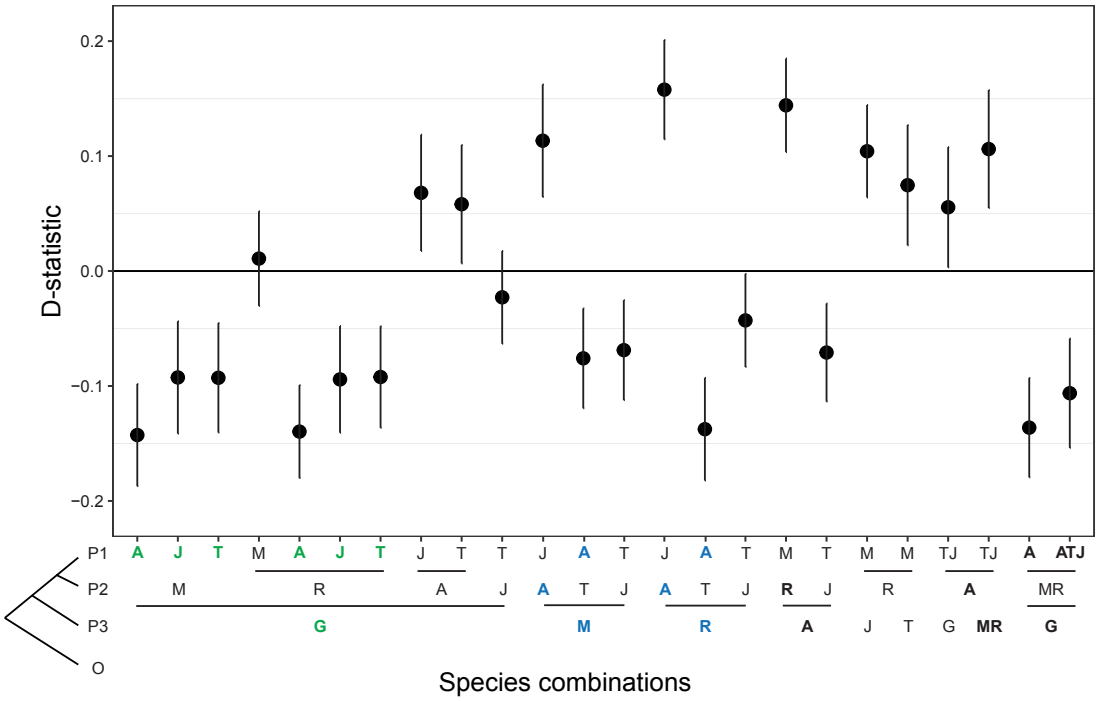
Appendix S5.7 Bar plots of STRUCTURE results for pairwise comparisons between species and lineages mainly distributed in the Top End.



Appendix S5.8 D statistics for species comparisons. T indicate approximate divergence distance from site counts (T12 – between P1 and P2, T123 – between P1 + P2 and P3). The column labeled 'Total SNPs' (from allele frequencies probabilities) gives the number of SNPs with ABBA and BABA patterns. Evidence for introgression was considered significant when the p-value was less than 0.01. A – *C. amax*, J – *C. johnstonei (s. l.)*, T – *C. triacantha (s. l.)*, G – *C. gracilis*, M – *C. munda*, R – *C. rufilatus*. Four additional tests lumping sister taxa were carried out to verify if results were consistent with those from the initial 20 tests.

P1	P2	P3	Total SNPs	T12	T123	D' stat	χ2	P-value	Introgression
A	M	G	1784	0.087	0.103	-0.161	46.493	0.000	A - G
J	M	G	1746	0.093	0.113	-0.109	20.676	0.000	J - G
T	M	G	1806	0.091	0.108	-0.118	25.358	0.000	T - G
M	R	G	1167	0.079	0.121	-0.015	0.248	0.619	-
A	R	G	1722	0.096	0.114	-0.174	52.265	0.000	A - G
J	R	G	1695	0.105	0.127	-0.127	27.271	0.000	J - G
T	R	G	1758	0.103	0.120	-0.137	32.765	0.000	T - G
J	A	G	1277	0.088	0.114	0.065	5.395	0.020	-
T	A	G	1320	0.086	0.109	0.050	3.300	0.069	-
T	J	G	754	0.073	0.129	-0.029	0.642	0.423	-
J	A	M	1247	0.075	0.087	0.113	15.943	0.000	A - M
A	T	M	1254	0.075	0.085	-0.097	11.869	0.001	A - M
T	J	M	723	0.062	0.097	-0.032	0.732	0.392	-
J	A	R	1347	0.082	0.097	0.145	28.229	0.000	A - R
A	T	R	1349	0.081	0.093	-0.136	24.825	0.000	A - R
T	J	R	761	0.069	0.109	-0.022	0.380	0.538	-
M	R	A	1212	0.066	0.091	0.116	16.172	0.000	R - A
T	J	A	795	0.062	0.090	-0.082	5.314	0.021	-
M	R	J	1089	0.074	0.097	0.089	8.640	0.003	R - J
M	R	T	1132	0.069	0.096	0.076	6.534	0.011	-
AJT	MR	G	1306	0.052	0.062	-0.133	23.182	0.000	AJT - G
A	MR	G	1390	0.071	0.085	-0.165	38.058	0.000	A - G
JT	A	G	1033	0.070	0.090	0.063	4.090	0.043	-
JT	A	MR	1009	0.054	0.062	0.110	12.211	0.000	A - MR

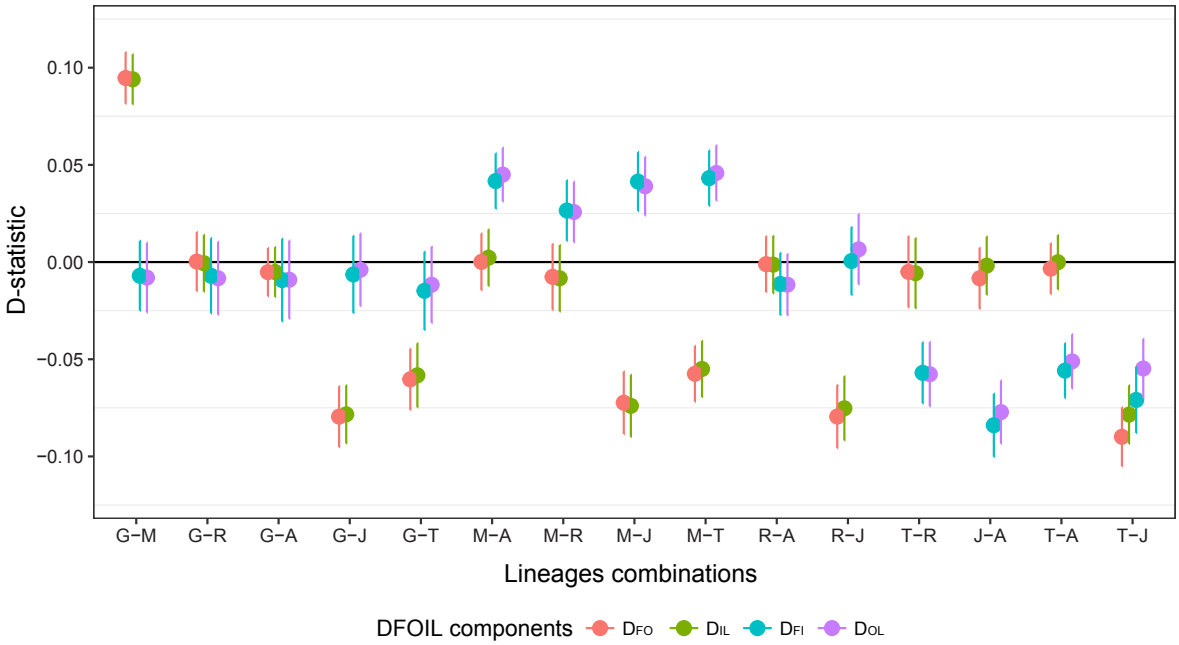
Appendix S5.9 95% confidence intervals from 1000 bootstraps for the D-statistic in the 20 species-level tests and the four additional ones with sister species lumped together in different combinations. Green correspond to tests supporting evidence for G-AJT introgression and blue for A-MR, while bold represents species with significant D-statistic.



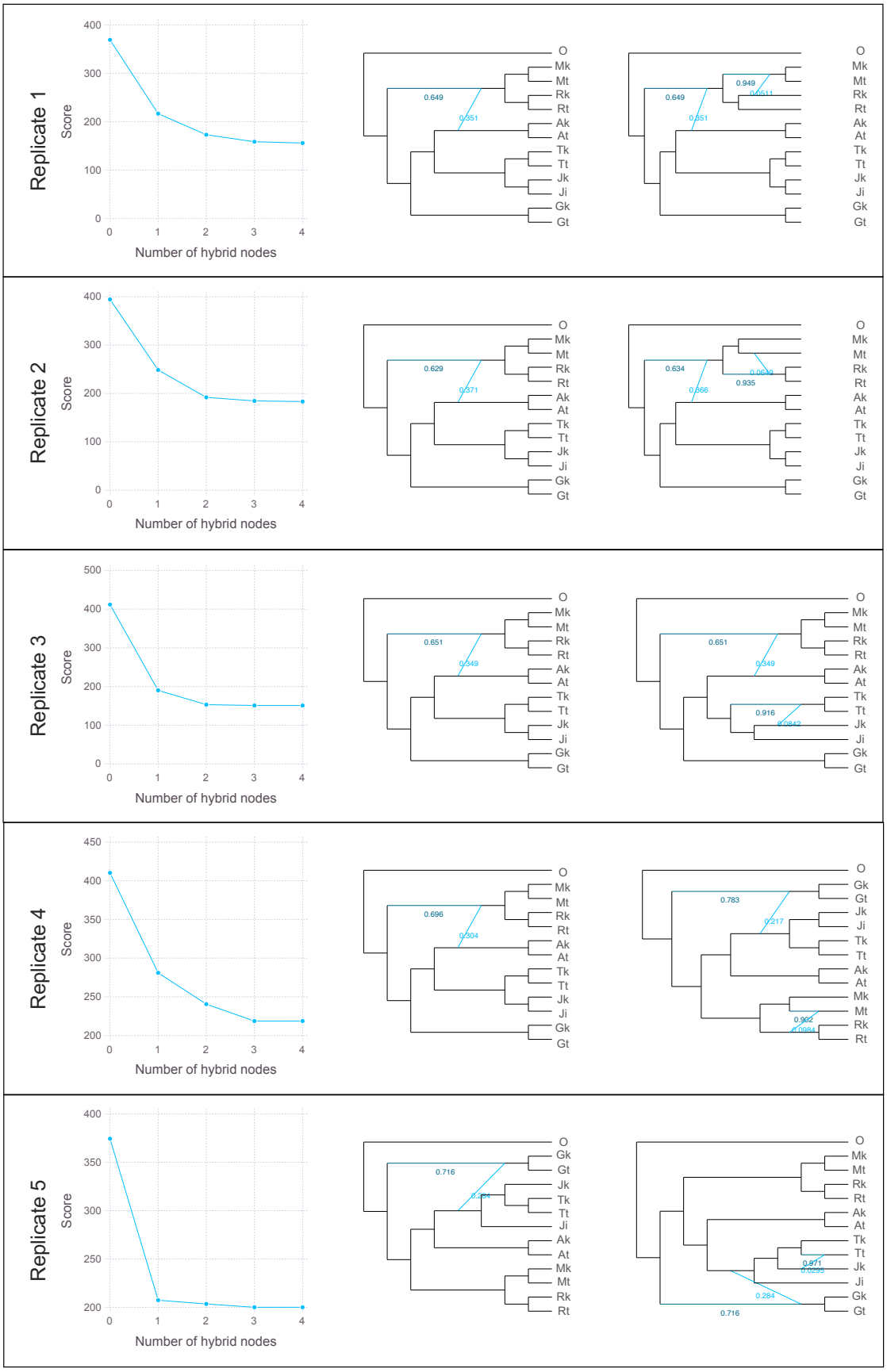
Appendix S5.10 Details of the D_{FOIL} results for species comparisons at the lineage level. For each of the four D-like statistics, information is given on the number of SNPs used (from allele frequencies probabilities), the value of the D_{FOIL} statistic component calculated as in Pease and Hahn 2015, the chi-square statistic and the respective p-value. Evidence, or lack thereof, for introgression is inferred from the combined results of the four test components, and the conclusions for each of the 15 comparisons are given in the introgression column.

P1	P2	P3	P4	Total SNPs	T12	T34	T1234	D_{FO}				D_{IL}				D_{FI}				D_{OL}				Introgression
								SNPs	D' stat	X^2	P-value	SNPs	D' stat	X^2	P-value	SNPs	D' stat	X^2	P-value	SNPs	D' stat	X^2	P-value	
Gk	Gt	Mk	Mt	10443	0.040	0.054	0.163	6167	0.085	44.354	0.000	6167	0.085	44.694	0.000	4421	-0.019	1.558	0.212	4421	-0.018	1.484	0.223	G - Mk
Gk	Gt	Rk	Rt	10966	0.047	0.069	0.199	6595	0.002	0.034	0.853	6595	0.003	0.044	0.834	4516	-0.016	1.148	0.284	4516	-0.016	1.085	0.298	-
Gk	Gt	Ak	At	13715	0.038	0.082	0.160	9526	-0.001	0.021	0.886	9526	-0.002	0.051	0.822	4407	-0.019	1.563	0.211	4407	-0.021	1.879	0.170	-
Gk	Gt	Jk	Ji	12780	0.048	0.096	0.186	8683	-0.080	54.990	0.000	8683	-0.081	56.593	0.000	4270	-0.010	0.453	0.501	4270	-0.013	0.683	0.409	G - Ji
Gk	Gt	Tk	Tt	10571	0.042	0.060	0.174	6345	-0.053	17.899	0.000	6345	-0.052	17.477	0.000	4370	-0.021	1.937	0.164	4370	-0.020	1.772	0.183	G - Tt
Mk	Mt	Ak	At	14949	0.041	0.066	0.113	9260	0.005	0.190	0.663	9260	0.005	0.249	0.618	5996	0.038	8.670	0.003	5996	0.039	9.132	0.003	-
Mk	Mt	Rk	Rt	12352	0.051	0.054	0.107	6455	0.004	0.082	0.775	6455	0.003	0.056	0.813	6137	0.021	2.712	0.100	6137	0.020	2.546	0.111	-
Mk	Mt	Jk	Ji	14024	0.049	0.072	0.124	8474	-0.077	50.783	0.000	8474	-0.078	52.029	0.000	5808	0.036	7.449	0.006	5808	0.034	6.887	0.009	-
Mk	Mt	Tk	Tt	11978	0.045	0.047	0.121	6191	-0.052	16.852	0.000	6191	-0.051	16.231	0.000	5984	0.038	8.840	0.003	5984	0.039	9.307	0.002	-
Rk	Rt	Ak	At	15612	0.051	0.075	0.131	9471	0.000	0.001	0.975	9471	0.002	0.024	0.878	6447	-0.004	0.113	0.737	6447	-0.002	0.035	0.852	-
Rk	Rt	Jk	Ji	14584	0.062	0.086	0.150	8657	-0.082	57.739	0.000	8657	-0.075	48.955	0.000	6209	0.005	0.175	0.675	6209	0.014	1.276	0.259	R - Ji
Tk	Tt	Rk	Rt	12476	0.055	0.055	0.143	6377	0.006	0.215	0.643	6377	0.005	0.132	0.716	6307	-0.051	16.338	0.000	6307	-0.052	17.162	0.000	-
Jk	Ji	Ak	At	17047	0.071	0.076	0.115	8992	-0.007	0.456	0.500	8992	-0.002	0.028	0.866	8465	-0.081	55.755	0.000	8465	-0.075	48.236	0.000	-
Tk	Tt	Ak	At	15102	0.047	0.070	0.112	9250	0.001	0.011	0.917	9250	0.001	0.004	0.950	6186	-0.046	13.223	0.000	6186	-0.047	13.595	0.000	-
Tk	Tt	Jk	Ji	13863	0.058	0.079	0.094	8442	-0.091	69.504	0.000	8442	-0.078	51.912	0.000	5933	-0.066	25.768	0.000	5933	-0.048	13.883	0.000	-

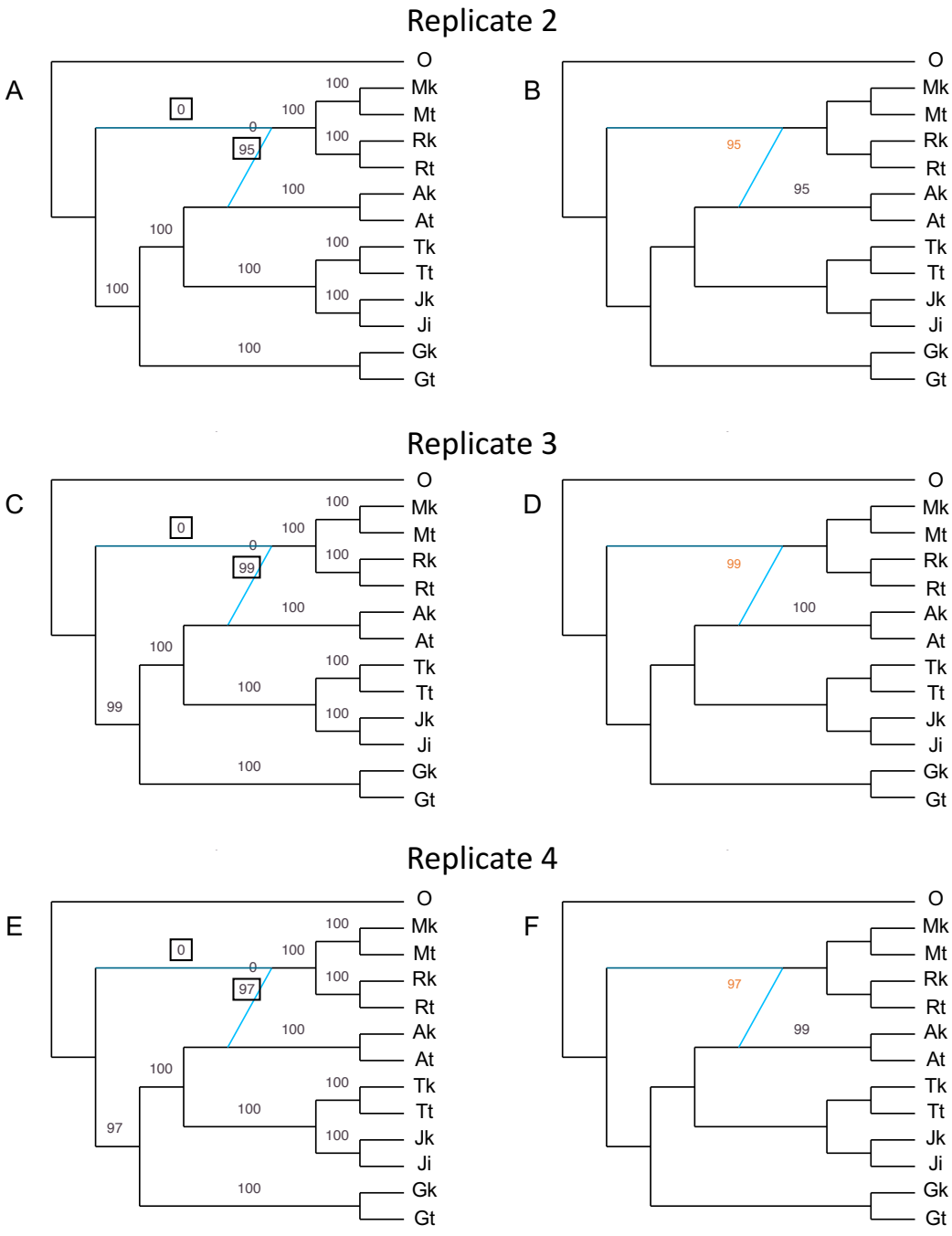
Appendix S5.11 95% confidence intervals from 1000 bootstraps for the four D_{FOIL} statistics in the pairwise species comparisons with Kimberley and Top End lineages. The details of the D_{FOIL} tests and conclusions regarding introgression for each of these comparisons are given in Table 5.1 and Appendix S5.10.



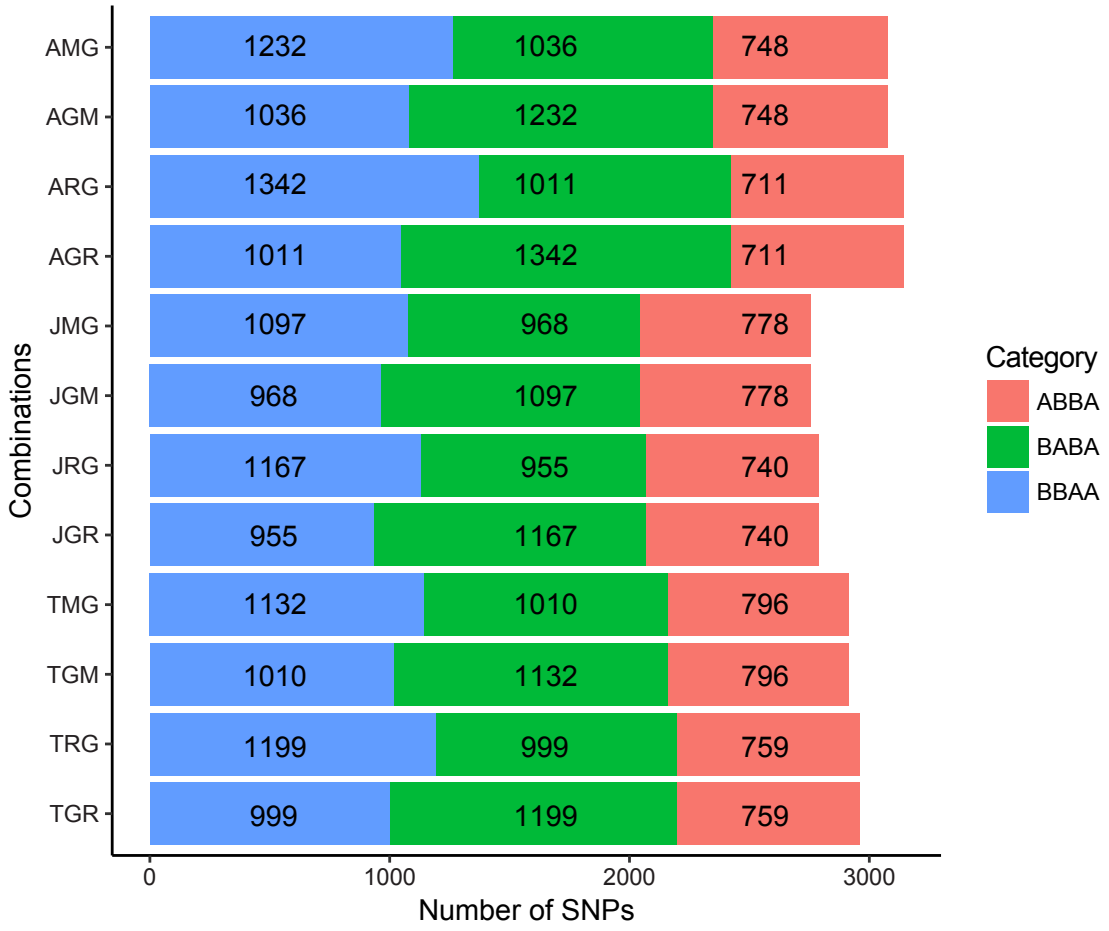
Appendix S5.12 Results of the five replicates of the PhyloNetworks analyses, showing the most likely number of reticulations and the best networks with, respectively, one and two reticulation events. In blue is the inheritance probability (γ), which represents the proportion of genes contributed by each parental population to a given hybrid node. Note that, for each replicate, networks with three and four reticulations were identical to those presented, but the second and third reticulation had γ of less than 10%. The networks with two reticulations do not recover both G-TJ and A-MR. The outlier replicate 5 shows that, although the A-MR reticulation was the most often supported (i.e. best pseudolikelihood), the G-TJ reticulation could have comparable pseudolikelihood, albeit across fewer runs.


Replicate 1
Score
Number of hybrid nodes
Replicate 2
Score
Number of hybrid nodes
Replicate 3
Score
Number of hybrid nodes
Replicate 4
Score
Number of hybrid nodes
Replicate 5
Score
Number of hybrid nodes
O
Mk
Mt
Rk
Rt
Ak
At
Tk
Tt
Jk
Ji
Gk
Gt

Appendix S5.13 Phylonetworks bootstrapped networks for replicates 2 (A and B), 3 (C and D) and 4 (E and F). The left panels show bootstrap values for branches; inside the squares are the bootstrap support for the hybrid edge (i.e. percentage of bootstrap trees in which the same sister clade is grouped with the hybrid clade) and the hybrid node. The right panels give the bootstrap values depending on whether a given lineage is the origin (above the edge) or target of gene flow (below the edge and in orange).



Appendix S5.14 Comparison of ABBA-BABA patterns across species combinations with *C. gracilis* as P3 or as P2 (i.e. considering respectively the two alternative topologies suggested by the phylogenetic network analyses). The expectation is that with the correct topology there will be more SNPs showing the BBAA pattern than the BABA pattern. The order in the species combinations is P1-P2-P3.



Chapter VI

General Discussion

The main purpose of this thesis was to understand how current species diversity patterns arose due to the effect of past climatic oscillations affecting speciation, range shifting and reticulation. To this end, I used a combination of next generation sequencing data and recent coalescent-based analyses to investigate these processes in rainbow skinks from the Australian Monsoonal Tropics (AMT), a group of closely related species that are sympatric in an understudied Australian biome. The main results of this study consist in the discovery of cryptic diversity within two species, with subsequent validation and description of new species, and finding genetic patterns in a climatic specialist that are consistent with it being more sensitive to past climatic change than a closely related sympatric generalist species. Additionally, I contributed to the generation of a robust phylogeny of the rainbow skinks that revealed close relationships between the six species that occur more broadly in the AMT, and identified several instances of evidence for ancient introgression despite negligible recent admixture.

6.1 Cryptic lineage diversity and phylogenetic relationships of *Carlia*

The discovery of cryptic species (two or more distinct species that are erroneously classified and hidden under one species name, Bickford *et al*. 2007) has exponentially increased with the development of genetic markers (Struck *et al.* 2018). Despite the description of such common patterns, the evolutionary processes behind this cryptic diversity are poorly studied in contrast to the long-studied mechanisms that produce morphological diversification (Fiser *et al.* 2017; Struck *et al.* 2018). Even the description of these patterns has often been done without any morphological analyses, using underpowered data with conflicting analyses under different assumptions (Fiser *et al.* 2017).

The description of overall patterns of cryptic diversity from a given region is the starting point for investigating hypothesis-driven questions to better understand those processes, and recent efforts have been done to identify this diversity in the understudied Australian Monsoonal Tropics (AMT) region (Oliver *et al.* 2012, 2016; Criscione & Köhler 2013; Marin *et al.* 2013; Huey *et al.* 2014; Moritz *et al.* 2016, 2018, Laver *et al.* 2017, 2018; Potter *et al.* 2018). This despite an already well-known morphospecies (species described through traditional taxonomy) diversity described for this region, including for the rainbow skinks (Storr 1974). For this relatively remote region, at the start of this study there was a generally sparse knowledge of species distributions, and lack of suitable specimens for genetic analysis (Moritz *et al.* 2013). Through extensive field collecting across several years, we have largely addressed this problem, accumulating for example for the AMT *Carlia* more than 900 tissue samples and close to 430 vouchered specimens. Nonetheless, some important sampling gaps remain for a few species (e.g. central Northern Territory for *C. triacantha* and *C. isostriacantha*).

Across several species in the genus *Carlia*, cryptic taxa have been identified within the Australian Wet Tropic, northeast Australia, Papua New Guinea (Stuart-Fox *et al.* 2002; Donnellan *et al.* 2009; Hoskin & Couper 2012; Rittmeyer 2014), and more recently in the

AMT (this study and Potter *et al.* 2016, 2017). Specifically, the latter studies, in six sympatric species (*C. munda, C. rufilatus, C. gracilis, C. johnstonei (s. l.), C. triacantha (s. l.),* and *C. amax*) identified 19 lineages, broadly overlapping among species and parapatric within species, with different range sizes across the Kimberley, Top End, Pilbara and Gulf regions. This suggests interesting dynamics in the AMT region that reflect the climatic fluctuations and the effect of aridity expansion in the region (further discussed bellow).

In this study, four lineages were identified in a pair of sister species, *C. johnstonei* and *C. triacantha*, and subsequent analyses, using morphological data and multispecies coalescent models, validated the lineages as new species. This second step is an important one that phylogeographers often do not undertake following identification of cryptic diversity, and this can have implications in fields that use species as units of analysis such as macroevolution or conservation (Fiser *et al.* 2017). One of the new species is *C. insularis*, a highly divergent species from *C. johnstonei* endemic to the small outermost laterite islands of the Kimberley. As the only vertebrate species endemic to these islands, *C. insularis* serves to highlight the great biodiversity value of the fauna of this region, which is considered an important conservation refuge (Gibson *et al.* 2017). The other new species is *C. isostriacantha*, which has most of it known distribution in the Kimberley but also extending to near the Gulf region. However, the previously mentioned sampling gap did not allow to define well its species boundaries with *C. triacantha* where they overlap. One individual with cytonuclear discordance was detected near this overlap region, but otherwise no evidence for current gene flow was observed between the two species. As with all species, *C. isostriacantha* was described as a hypothesis that might change with more sampling and data.

Some proposed mechanisms causing cryptic diversity are recent divergence, phylogenetic niche conservatism, stasis, parallelism and convergence (Fiser *et al.* 2017; Struck *et al.* 2018). A well-defined resolution of phylogenetic relationships between species is required to better understand the processes that might have caused the cryptic diversity observed in *Carlia* species. The estimate of phylogenetic relationships by Dolman & Hugall (2008, Fig. 1.1 in Chapter I) showed much uncertainty between species within *Carlia* and with other related genera in the same clade. In this thesis, I participated in a study producing a much more robust phylogeny for the relationships between species of the three genera of rainbow skinks (*Carlia*, *Lygisaurus* and *Liburnascincus*). At variance with Dolman & Hugall (2008), this study showed that the six AMT species are actually closely related, and, surprisingly, that their clade includes *C. tetradactyla* from the temperate southeast of Australia. The differences between these studies show how increasing the number of analysed loci, in this case by three orders of magnitude, can help to infer accurate relationships, especially in fast radiations like *Carlia*. Moreover, although it seems that mtDNA alone is often sufficient to identify the distribution of cryptic lineages within the AMT *Carlia* (Chapter II, Potter *et al.* 2017), it fails to resolve relationships between taxa (see also Stuart-Fox *et al.* 2002).

The species of *Carlia* display many similar traits, with some differences between species in their breeding colours, ear lobules, dorsal scales characteristics and in a few head scales

(Cogger 2014). The morphological variability across species seems to overlap between different groups of species (Zug 2010), but no strong convergence seems to be in play, as in the case of the *Cryptoblepharus* skink genus where there is ecomorphological convergence with cryptic diversification (Blom *et al.* 2016). Recent divergence also seems an unlikely process because several cryptic lineage splits are deeper than within other described species (e.g. C. pectoralis in Chapter IV). To further test for the processes behind cryptic diversity, more phenotypic and ecological data needs to be collected across the genus *Carlia*, or even more broadly including the Eugongylus group, since several of its clades also show cryptic diversity (Blom *et al.* 2016, Moritz & Potter personal communication).

6.2 Responses of species with different climatic niches to past climatic change

Realized niches of species, as inferred by intersecting distribution records with climate layers, can provide insight to their responses to past climate change (Nogués-Bravo 2009). Moreover, differences in niche breadth may suggest differences in sensitivity between species (Thuiller *et al.* 2005). Thus, studying species niches can be particularly relevant in lizards to understand how species may have responded to past climate change, and they are also a great system to study speciation and divergence processes (Camargo *et al.* 2010). This is because lizards can be very sensitive to climatic fluctuations due to their behaviour, low dispersal and physiological capabilities (Huey *et al.* 2009).

Although we intuitively expect that species with broadly similar realised climatic niches would have responded similarly to past environmental changes, such species can also show contrasting responses due to differences in key traits that may only be revealed with the study of phylogeographic patterns (Zamudio *et al.* 2016; Prates *et al.* 2016), and that may be related to ecological, demographic or phenotypic processes, like differences in habitat requirements or in dispersal rates. Similar historical responses were identified across four *Carlia* species from the Top End region, with similar biogeographic breaks and evidence for synchronous expansion in a region with less climatic instability during the LGM (Potter *et al.* 2017). By comparison, across the region with more climatic instability, the Kimberley, the same set of species showed more variable responses, suggesting that trait-based differences in sensitivity manifest more strongly in a harsher, more variable environment.

Taking into account the revised phylogeny for AMT *Carlia*, two comparisons of sister taxa with different climatic breadth can be made: *C. munda* vs *C. rufilatus* and *C. johnstonei (s. l.)* vs *C. triacantha (s. l.)*. Both *C. munda* and *C. triacantha (s. l.)* have a broad distribution across the tropical north and arid centre of Australia, and both show modest phylogeographic structure, each with geographically widespread lineages. However, these taxa have different histories, as *C. munda* shows a more tropical (Top End) to subtropical-arid split, while the second taxon shows a less clear pattern, considering the split between the subtropical *C. isostriacantha* and the tropical to arid *C. triacantha (s. s.)*. A dated-tree shows that the split in both taxa may have occurred at about the same time (from an ongoing study; Fig. 6.1, Supplementary Material). Each of the more climatically restricted taxa, *C. rufilatus*

and *C. johnstonei (s. l.)*, has deeply divergent island lineages (*C. rufilatus* ECI in the Top End and *C. insularis* in the north-west Kimberley) that are associated with wetter environments. By contrast, the broader-niche species show less evidence of strong divergence between island and mainland populations. The divergence of both island lineages (Fig. 6.1, Supplementary Material) in the mesic-specialist taxa might have occurred around a period of sharp transition from arid to an increasingly humid climate in northern Australia around 5.5 MYA (Christensen *et al.* 2017). Notably, the period in the late Pliocene (around 2.7 MYA) when the mesic-arid splits occurred within the generalist species was around the same to when the Kimberley and Top End lineages of the more climatic specialist *C. rufilatus* diverged, but for this species an expansion to a more arid environment is not observable (Fig. 6.1, Supplementary Material). Interestingly, these splits are consistent with the period of transition to a more arid and seasonal climate that was set around 2.4 MYA (Christensen *et al.* 2017). Although paleo-SDM models cannot predict species distributions when these Pliocene splits occurred, for *C. rufilatus* at the LGM the models showed that the rainfall in the Kimberley was lower than the realized climatic niche of the species (Potter *et al.* 2018), while the *C. johnstonei* model could not predict the occurrence of the species (unpublished data). This difficulty to model past responses of these species suggests the need of more mechanistic models for climatic specialist species that were more sensitive to past climatic fluctuations.

Intrinsic species' traits could explain the concordant and discordant phylogeographic patterns between these taxa, but for the AMT species almost no life-history (clutch size, James & Shine 1988) or ecological information is published. My unpublished data from field-based physiological experiments suggest an overlap of thermal limits across species, except for *C. gracilis* that had higher critical temperature minimum (CT_{min}) and lower CT_{max} than the other five species in the Kimberley. Also, my unpublished analyses of records from standardised ecological surveys in the north Kimberley (C. Myers, pers. comm.) suggest some differences in habitat use: *C. gracilis* is more associated with riparian and wetland habitats, *C. munda* with savanna grassland, *C. isostriacantha* with sandstone, and *C. johnstonei* with volcanic rocks. These preliminary data for ecological traits seem to be suggesting that although overlapping lineages might be capable of enduring similar climatic fluctuations, the differences in habitat preferences in a heterogeneous environment could have contributed to disparate responses of species to past climate fluctuations (Zamudio *et al.* 2016). Clearly, further ecological and physiological data are needed to enable more direct comparisons, e.g. using mechanistic models and habitat-adjusted climate surfaces (Kearney *et al.* 2010, 2014; Kearney & Porter 2016; Fordham *et al.* 2016).

A better understanding of the ecology of these species would also help with formulating more informed hypotheses regarding past hybridization events. For example, a testable hypothesis is that a difference in niche breadth of introgressed taxa (e.g. *C. munda* BROAD and the restricted *C. insularis*) is related to a difference in effective population size and/or potential historical expansion of one species but not the other, with potential consequences resulting in hybridization (Excoffier *et al.* 2009). We can employ ABC (Shafer *et al.* 2015) to

compare these competing demographic history models, by estimating effective population sizes while also estimating both the direction and magnitude of gene flow. Additionally, it might be possible to estimate if the timing of any historical expansions or gene flow between these taxa coincides with the known shift to an arid climate in the north of Australia (Christensen *et al.* 2017). However, interpreting events in the deep past, such as inferred in Chapter V, would require estimating ancestral traits, which could be unlikely given the very different ecological requirements of extant species.

6.3 Ancestral introgression between Australian Monsoonal Tropics species

The rise in studies using genomic data has led to increased evidence of gene flow during speciation and also introgression between divergent taxa, expanding the detection of historical hybridization beyond finding cytonuclear discordances (Abbott *et al.* 2013; Bonnet *et al.* 2017). Likewise, methodological advancements have allowed researchers to explore further patterns of introgression and to better understand the effect of gene flow in diversification and in the genomics of speciation (Abbott *et al.* 2013; Payseur & Rieseberg 2016).

Exploring these patterns in closely related sympatric species increases the probability of finding patterns of current hybridization or of past introgression. In cryptic species or species that are morphologically very conserved, testing the occurrence of such patterns based only on morphology may be unfeasible. Similarly, relating introgression patterns with phenotypic traits may also prove hard in cases involving cryptic species (e.g. Chattopadhyay *et al.* 2016). Thus, an increase in the number of studies testing for introgression in sympatric pseudo-cryptic species (i.e. species that can only be identified by detailed comparison of morphological and non-morphological features; Sáez & Lozano 2005) may begin to show that introgression is more common than expected between species where morphological hybrids are not easy to detect.

This study, which mostly intended to understand differences in introgression at regional scales, unexpectedly provided evidence for ancient introgression between *Carlia* species from the AMT, but little, if any, over recent generations. The several ancient introgression events detected are likely to be in the order of several million years old. This is suggested by a preliminary dated tree from an ongoing study (Fig. 6.1, Supplementary Material) that shows when the four more recent events likely occurred. Considering that the Kimberley – Top End *C. gracilis* divergence is very recent (~1 MYA), the three inferred ancient hybridization events in which it was involved may be the more recent, while the *C. rufilatus* – *C. insularis* introgression could have occurred anytime between 5 to 2.7 MYA. However, it is important to note that this dated tree may be biased because MSC models underestimate branch lengths among taxa exchanging genes, even if the topology is robust to the presence of gene flow (Heled *et al.* 2013; Leaché *et al.* 2014; Wen & Nakhleh 2017).

The deeper introgression events seem to have occurred early in the history of this rapid radiation, considering the very short nodes between *C. gracilis*, the *C. munda* + *C. rufilatus*

ancestor, *C. tetradactyla* (the southeastern species) and the remaining species (Fig. 6.1, Supplementary Material). Assuming the scenario with two deeper introgressions, G-AJT and A-MR, it seems likely that these events might have happened around the time when species were diverging (8 to 9 MYA and 7 to 8 MYA, respectively; Fig. 6.1, Supplementary Material), and if the ancestral taxa were sympatric, as the descendant species currently are, the divergence might have therefore occurred in the presence of gene flow. In turn, this suggests that *Carlia* could be another example of an association between rapid radiations and introgression, perhaps analogous to cichlid fishes from African lakes (Gante *et al.* 2016; Meier *et al.* 2017), *Heliconius* butterflies (*Heliconius* Consortium 2012) or Darwin's finches (Grant & Grant 2014), but now illustrating that such processes do not necessarily have to be accompanied by strong phenotypic evolution. It is possible that more studies investigating ancient introgressive hybridization across nuclear genomes in a radiation of cryptic species or in groups with low phenotypic differentiation, might find this to be a more common pattern, probably facilitated by low genetic divergence during rapid radiation.

Introgression can also distort phylogenetic analyses that assume purely divergent evolution (Solís-Lemus *et al.* 2016). The unexpected position of *C. gracilis* in this case study could indeed suggest that species tree methods may not be able to recover the correct topology if presence of gene flow is an assumption violation. However, the topology from StarBEAST2, a method expected to be more robust to the effect of gene flow, is consistent with that from the other species tree methods. This conflicting result underlines the need of further studies to verify the effect of gene flow in fast radiations when inferring introgression with phylogenetic network methods. These are recent methods that need to be more extensively tested for effects of model violations. My results should be also assessed using datasets other than exons, like whole-genome data. This could show whether exon data may be more favourable to detect introgression, identify which genomic regions show more evidence of introgression, test if these regions are under selection, and shed light on how the genomic landscape of introgression relates with speciation (Martin & Jiggins 2017; Ravinet *et al.* 2017).

6.4 Application of exon capture data in phylogenetics and phylogeography

This study shows with several different questions the advantage that using 1000's of loci brings to investigate phylogenetics and phylogeography. Specifically, in taxa that went through rapid diversification and/or past hybridization, a large number of loci is necessary to identify and disentangle these patterns (Edwards *et al.* 2016). Exon capture was shown to be particularly helpful when studying questions across the speciation continuum (Bragg *et al.* 2016; Edwards *et al.* 2016; Potter *et al.* 2018), since it can provide sequence data useful to resolve species, their relationships and demographic histories (e.g. Duchêne *et al.* 2017; O'Hara *et al.* 2017; Moritz *et al.* 2018) and loci are still variable enough to be informative for population genetics (e.g. Schweizer *et al.* 2016; Singhal & Bi 2017).

Finding robust topological relationships using NGS data can nonetheless be a substantial challenge (Chou *et al.* 2015; Roch & Warnow 2015). Dating without calibration fossils can be also very uncertain (dos Reis & Yang 2013), but analysing multi-locus data across different scales and statistical methods can provide more confidence in estimates. In this study, I used a demographic method to obtain estimates for lineage divergence, which seem to be fairly well corroborated by estimates from a MSC method that has underlying different assumptions (Chapter II; Fig. 6.1, Supplementary Material). This was achieved retrieving SNPs, which were subsequently summarized in SFS for the demographic approach, while using sequence data loci for the coalescent-based method.

However, several hundreds to thousands of loci can be much more data than is computationally feasible for many MSC approaches. Thus decisions need to be considered to take full advantage of the data, like using multiple independent datasets, either in samples or in loci (Chapter III and IV), or using some criteria to select small sets of loci (Lemmon & Lemmon 2013; Blom *et al.* 2017).

Some caution with SNP-based analysis should also be reflected when choosing a random SNP per locus regarding datasets with many missing data, either by samples or by sites, with some suggestions reviewed in Leaché & Oaks (2017). A restriction on how much missing data can be considered for a site to be selected can be helpful (Blom 2015). The number of SNPs that could be considered with this exon capture dataset (around 1700 – 2000, depending on the analysis) can be insufficient for more complex demographic models. Parameter estimates across runs using these models were less consistent (data not shown), but further analysis with simulations are needed to demonstrate the limits to inferring demographic parameters from few thousand SNP loci.

For specific questions, whole-genome is becoming much more useful and the costs for non-model organisms are rapidly decreasing (Goodwin et al. 2016). That said, sequence capture datasets are a great first pass resource for the study of evolutionary and ecological genomics of any species (Jones & Good 2016). Only after disentangling cryptic species and their internal structure will whole-genome sequencing be more cost-effective, and likely of better use in hypothesis-driven studies.

6.5 Final considerations and future perspectives

Until very recently, phylogeographic studies had described more intraspecific diversity in Northern Hemisphere species (Beheregaray 2008), but recent efforts are better describing the high levels of biodiversity and patterns in the Southern Hemisphere, specifically in Australia that has a rich and unique Gondwanan fauna and flora (Byrne *et al.* 2008, 2011; Bowman *et al.* 2010; Rix *et al.* 2015; Edwards *et al.* 2017). This study is part of a major collaborative effort to increase the description of biodiversity and the understanding of evolutionary processes occurring in species from the rich savannah and relatively remote Australian Monsoonal Tropics biome.

Here I applied advanced methodologies to disentangle patterns related to speciation and hybridization in pseudo-cryptic skinks with unknown intraspecific cryptic diversity. Particularly, this study went further than most phylogeographic studies, in which taxonomical revisions are often not done (Carstens *et al.* 2013), by describing two new species from the identified cryptic lineages, and other species descriptions could be warranted for other highly divergent lineages within species of *Carlia*. As mentioned, species are hypothesis that should be tested with more data and models, so it would also be interesting to further test them with future expected approaches based on the protracted speciation model (as proposed in Knowles and Sukumaran 2017).

Much work has shown the effect of Quaternary climatic cycles in promoting hybridization (reviewed in Hewitt 2011), especially in temperate regions that were heavily glaciated. I present a case study where ancient hybridization was detected and possibly induced by climatic fluctuations due to arid-humid shifts, suggesting how changes in precipitation might be a strong factor like the thermal variations during glacial cycles, especially in regions not affected by ice sheets. This is also an example of how to detect ancient hybridization beyond studies of parapatric lineages with a current hybrid zone.

This study also shows how cryptic divergence within and across sympatric species groups might hide processes of hybridization, and serves as a starting point to ask how has introgression been part of the evolutionary processes in the rapid radiation of the Eugongylus group and, more generally, in cryptic complexes of tropical taxa.

Besides suggestions made throughout this discussion, future studies are needed to increase sampling, and ecological and behaviour knowledge for *Carlia* species and other AMT species, especially during the wet season when most species display breading season features. The remote access to some regions in the AMT, for example the Kimberley and nearby islands, make such studies costly and often hard due to absence of support facilities (Moritz *et al.* 2013; Rosauer *et al.* 2016, 2018). As mentioned previously, such studies are particularly important for the AMT herpetofauna since its diversity of morphologically similar species can be suggesting important evolutionary and ecological processes where sexual selection might also have a role, but so far, the study of these mechanisms is still lacking in theoretical and empirical research. This study and other ongoing work are the starting point for studying such macroevolutionary processes shaping phenotypic, ecological and speciation patterns.

In addition, the genetic patterns described in this study can suggest hypotheses to further explore the genomics of speciation using whole-genome sequencing, particularly to infer which genomic regions show more introgression, if these are under selective forces and how these processes relate to species capacity to adapt to changing climates.

6.6 References

Abbott R, Albach D, Ansell S et al. (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Beheregaray LB (2008) Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.

Bickford D, Lohman DJ, Sodhi NS et al. (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148–155.

Blom MPK (2015) EAPhy: A flexible tool for high-throughput quality filtering of exon-alignments and data processing for phylogenetic methods. *PLoS Currents Tree of Life*, **1**.

Blom MPK, Bragg JG, Potter S, Moritz C (2017) Accounting for uncertainty in gene tree estimation: Summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Systematic Biology*, **66**, 352–366.

Blom MPK, Horner P, Moritz C (2016) Convergence across a continent: adaptive diversification in a recent radiation of Australian lizards. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20160181.

Bonnet T, Leblois R, Rousset F, Crochet P-A (2017) A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, **71**, 2140–2158.

Bowman DMJS, Brown GK, Braby MF et al. (2010) Biogeography of the Australian monsoon tropics. *Journal of Biogeography*, **37**, 201–216.

Bragg JG, Potter S, Bi K, Moritz C (2016) Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources*, **16**, 1059–1068.

Byrne M, Steane D a., Joseph L et al. (2011) Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography*, **38**, 1635–1656.

Byrne M, Yeates DK, Joseph L et al. (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology*, **17**, 4398–417.

Camargo A, Sinervo B, Sites JW (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology*, **19**, 3250–70.

Carstens BC, Pelletier T A., Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Chattopadhyay B, Garg KM, Kumar AKV et al. (2016) Genome-wide data reveal cryptic diversity and genetic introgression in an Oriental cynopterine fruit bat radiation. *BMC Evolutionary Biology*, **16**, 41.

Chou J, Gupta A, Yaduvanshi S et al. (2015) A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, **16**, S2.

Christensen BA, Renema W, Henderiks J et al. (2017) Indonesian Throughflow drove Australian climate from humid Pliocene to arid Pleistocene. *Geophysical Research Letters*, **44**, 6914–6925.

Cogger H (2017) *Reptiles and amphibians of Australia*. New Holland, Sydney.

Criscione F, Köhler F (2013) Six new species of *Australocosmica* Köhler, 2011 from the Kimberley Islands, Western Australia (Mollusca: Pulmonata: Camaenidae). *Zootaxa*, **3608**, 101–15.

Dolman G, Hugall AF (2008) Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Molecular Phylogenetics and Evolution*, **49**, 782–94.

Donnellan S, Couper P, Saint K, Wheaton L (2009) Systematics of the *Carlia* "*fusca*" complex (Reptilia: Scincidae) from northern Australia. *Zootaxa*, **2227**, 1–31.

Duchêne DA, Bragg JG, Duchêne S et al. (2017) Analysis of Phylogenomic Tree Space Resolves Relationships Among Marsupial Families. *Systematic Biolo*gy, syx076

.

Edwards RD, Crisp MD, Cook DH, Cook LG (2017) Congruent biogeographical disjunctions at a continent-wide scale: Quantifying and clarifying the role of biogeographic barriers in the Australian tropics. *PLoS ONE*, **12**, 1–17.

Edwards S V., Potter S, Schmitt CJ, Bragg JG, Moritz C (2016) Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences*, **113**, 8025–8032.

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Fiser C, Robinson CT, Malard F (2017) Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, **12**, 3218–3221.

Fordham DA, Akçakaya HR, Alroy J et al. (2016) Predicting and mitigating future biodiversity loss using long-term ecological proxies. *Nature Climate Change*, **6**, 909.

Gante HF, Matschiner M, Malmstrøm M et al. (2016) Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Molecular Ecology*, **25**, 6143–6161.

Gibson LA, Cowan MA, Lyons MN et al. (2017) Island refuges: Conservation significance of the biodiversity patterns resulting from "natural" fragmentation. *Biological Conservation*, **212**, 349–356.

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17**, 333–351.

Grant PR, Grant BR (2014) Synergism of natural selection and introgression in the origin of a new species. *The American Naturalist*, **183**, 671–681.

Heled J, Bryant D, Drummond AJ (2013) Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evolutionary Biology*, **13**, 44.

Hewitt GM (2011) Quaternary phylogeography: The roots of hybrid zones. *Genetica*, **139**, 617–638.

Hoskin CJ, Couper PJ (2012) Description of two new *Carlia* species (Reptilia: Scincidae) from northeast Australia, elevation of *Carlia pectoralis inconnexa* Ingram & Covacevich 1989 to full species status, and redescription of *Carlia pectoralis* (de Vis 1884). *Zootaxa*, **28**, 1–28.

Huey JA, Cook BD, Unmack PJ, Hughes JM (2014) Broadscale phylogeographic structure of five freshwater fishes across the Australian Monsoonal Tropics. *Freshwater Science*, **33**, 273–287.

Huey RB, Deutsch CA, Tewksbury JJ et al. (2009) Why tropical forest lizards are vulnerable to climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 1939–48.

James C, Shine R (1988) Life-history strategies of Australain lizards: a comparison between the tropics and the temperate zone. *Oecologia*, **75**, 307–316.

Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, **25**, 185–202.

Kearney MR, Porter WP (2016) NicheMapR - an R package for biophysical modelling: the microclimate model. Ecography.

Kearney MR, Shamakhy A, Tingley R et al. (2014) Microclimate modelling at macro scales: a test of a general microclimate model integrated with gridded continental-scale soil and weather data. *Methods in Ecology and Evolution*, **5**, 273–286.

Kearney MR, Wintle B a., Porter WP (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203–213.

Laver RJ, Doughty P, Oliver PM (2018) Origins and patterns of endemic diversity in two specialized lizard lineages from the Australian Monsoonal Tropics (*Oedura* spp.). *Journal of Biogeography*, **45**, 142–153.

Laver RJ, Nielsen S V, Rosauer DF, Oliver PM (2017) Trans-biome diversity in Australian grass-specialist lizards (Diplodactylidae: Strophurus). *Molecular Phylogenetics and Evolution*, **115**, 62–70.

Leaché AD, Harris RB, Rannala B, Yang Z (2014) The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology*, **63**, 17–30.

Leaché AD, Oaks JR (2017) The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **48**, 69–84.

Lemmon EM, Lemmon AR (2013) High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Marin J, Donnellan SC, Hedges SB et al. (2013) Hidden species diversity of Australian burrowing snakes (*Ramphotyphlops*). *Biological Journal of the Linnean Society*, **110**, 427–441.

Martin SH, Jiggins CD (2017) Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development*, **47**, 69–74.

Meier JI, Marques DA, Mwaiko S et al. (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, **8**, 14363.

Moritz C, Ens EJ, Potter S, Catullo RA (2013) The Australian monsoonal tropics: An opportunity to protect unique biodiversity and secure benefits for Aboriginal communities. *Pacific Conservation*, **19**, 343–355.

Moritz C, Fujita MK, Rosauer D et al. (2016) Multilocus phylogeography reveals nested endemism in a gecko across the monsoonal tropics of Australia. *Molecular Ecology*, **25**, 1354–1366.

Moritz CC, Pratt RC, Bank S et al. (2018) Cryptic lineage diversity, body size divergence, and sympatry in a species complex of Australian lizards (*Gehyra*). *Evolution*, **72**, 54–66.

Nogués-Bravo D (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, **18**, 521–531.

O'Hara TD, Hugall AF, Thuy B, Stöhr S, Martynov A V (2017) Restructuring higher taxonomy using broad-scale phylogenomics: The living Ophiuroidea. *Molecular Phylogenetics and Evolution*, **107**, 415–430.

Oliver PM, Doughty P, Palmer R (2012) Hidden biodiversity in rare northern Australian vertebrates: The case of the clawless geckos (*Crenadactylus*, Diplodactylidae) of the Kimberley. *Wildlife Research*, **39**, 429–435.

Oliver PM, Laver RJ, De Mello Martins F et al. (2016) A novel hotspot of vertebrate endemism and an evolutionary refugium in tropical Australia. *Diversity and Distributions*, **23**, 53–66.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337–2360.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, **25**, 1367–1380.

Potter S, Xue AT, Bragg JG et al. (2017) Pleistocene climatic changes drive diversification across a tropical savanna. *Molecular Ecology*. In Press. doi: 10.1111/mec.14441.

Prates I, Xue AT, Brown JL et al. (2016) Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences*, **113**, 7978–7985.

Ravinet M, Faria R, Butlin RK et al. (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, **30**, 1450–1477.

dos Reis M, Yang Z (2013) The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematics and Evolution*, **51**, 30–43.

Rittmeyer EN (2014) *Systematics and Species Delimitation in New Guinea Skink Species Complexes (Squamata: Scincidae)*. LSU Doctoral Dissertations. 287.

Rix MG, Edwards DL, Byrne M et al. (2015) Biogeography and speciation of terrestrial fauna in the south-western Australian biodiversity hotspot. *Biological Reviews*, **90**, 762–793.

Roch S, Warnow T (2015) On the Robustness to Gene Tree Estimation Error (or lack thereof) of Coalescent-Based Species Tree Methods. *Systematic Biology*, **64**, 663–676.

Rosauer DF, Blom MPK, Bourke G et al. (2016) Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia. *Biological Conservation*, **204**, 83–93.

Rosauer DF, Byrne M, Blom MPK et al. (2018) Real-world conservation planning for evolutionary diversity in the Kimberley, Australia, sidesteps uncertain taxonomy. *Conservation Letters*, e12438.

Sáez AG, Lozano E (2005) Body doubles. *Nature*, **433**, 111.

Schweizer RM, Robinson J, Harrigan R et al. (2016) Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Molecular Ecology*, **25**, 357–379.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW (2015) Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. *Molecular Ecology*, **24**, 328–345.

Singhal S, Bi K (2017) History cleans up messes: The impact of time in driving divergence and introgression in a tropical suture zone. *Evolution*, **71**, 1888–1899.

Solís-Lemus C, Yang M, Ané C (2016) Inconsistency of Species Tree Methods under Gene Flow. *Systematic Biology*, **65**, 843–851.

Storr GM (1974) The genus *Carlia* (Lacertilia, Scincidae) in Western Australia and Northern Territory. *Records of the Western Australian Museum*, **3**, 151–165.

Struck TH, Feder JL, Bendiksby M et al. (2018) Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology & Evolution.* **33**, 153 -163.

Stuart-Fox DM, Hugall AF, Moritz C (2002) A molecular phylogeny of rainbow skinks (Scincidae: *Carlia*): taxonomic and biogeographic implications. *Australian Journal of Zoology*, **50**, 39–51.

The *Heliconius* Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Thuiller W, Lavorel S, Araújo MB (2005) Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology and Biogeography*, **14**, 347–357.

Wen D, Nakhleh L (2017) Coestimating Reticulate Phylogenies and Gene Trees from Multilocus Sequence Data. *Systematic Biology*, syx085.

Zamudio KR, Bell RC, Mason NA (2016) Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences*, **113**, 8041–8048.

Zug GR (2010) An Outlying Carlia Population from Java and Comments on Species Groups within the Genus *Carlia* (Reptilia: Squamata: Scincidae). *Proceedings of the California Academy of Sciences*, **61**, 389–408.

6.7 Supplementary Material

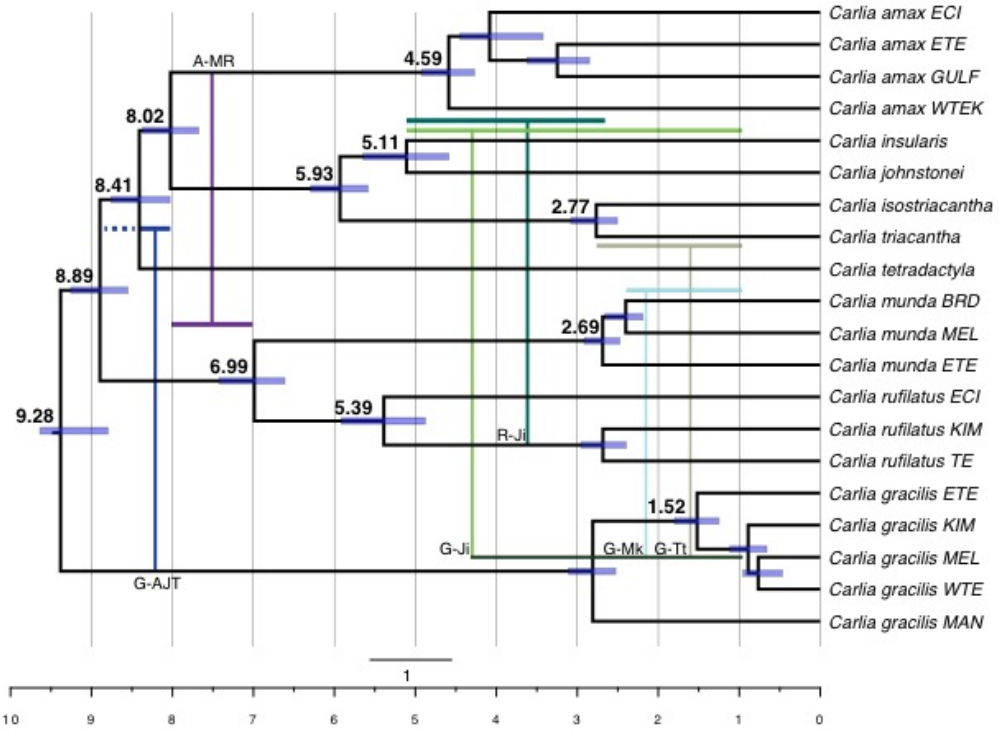


Figure 6.1 Preliminary time-tree with confidence intervals from ongoing collaborative study aiming to investigate macroevolutionary patterns across the Eugongylus group. This is part of a StarBeast2 tree produced with 100 random loci and a strict clock, using a substitution rate similar to what was used in the demographic analysis of Chapter II (10^{-9} substitutions per generation). This tree was estimated with more species, but here are only shown the relationships between the lineages studied in this study (including the position for *C. tetradactyla*). The estimates for divergence times between *C. insularis - C. johnstonei* and between *C. triacantha – C. isostriacantha* are very close to the demographic estimates obtained in Chapter II (~5.6 and 2.2, respectively). The different colours correspond to potential intervals of time, corresponding to the ancestral branches involved, in inferred introgression events (designation of events as in Chapter V).